FRONTIERS IN PHYSICS

Modern Kaluza-Klein Theories

Thomas Appelquist
Alan Chodos
Peter G.O. Freund

~~FALVEY MEMORIAL LIBRARY~~
~~VILLANOVA UNIVERSITY~~
~~VILLANOVA, PA. 19085~~

NO LONGER PROPERTY OF
FALVEY MEMORIAL LIBRARY

MODERN KALUZA-KLEIN THEORIES

Sponsoring Editor: Allan Wylde
Production Supervisor: Karen Gulliver
Production: Editing, Design & Production, Inc.

Copyright © 1987 by Addison-Wesley Publishing Company, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America. Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Modern Kaluza-Klein theories.

(Frontiers in physics; v. 65)
1. Kaluza-Klein theories. I. Appelquist, Thomas.
II. Chodos, Alan. III. Freund, Peter G. O. (Peter George Oliver), 1936- IV. Series.
QC793.3.F5M63 1987 530.1'4 86-26613
ISBN 0-201-09829-6

ABCDEFGHIJ-MA-89876

FRONTIERS IN PHYSICS

David Pines, Editor

Volumes of the Series published from 1961 to 1973 are not officially numbered. The parenthetical numbers shown are designed to aid librarians and bibliographers to to check the completeness of their holdings.

Titles published in this series prior to 1987 appear under either the W.A. Benjamin or the Benjamin/Cummings imprint; titles published since 1986 appear under the Addison-Wesley imprint.

(1)	N. Bloembergen	Nuclear Magnetic Relaxation: A Reprint Volume, 1961
(2)	G. F. Chew	*S*-Matrix Theory of Strong Interactions: A Lecture Note and Reprint Volume, 1961
(3)	R. P. Feynman	Quantum Electrodynamics: A Lecture Note and Reprint Volume, 1961 (8th printing, 1983).
(4)	R. P. Feynman	The Theory of Fundamental Processes: A Lecture Note Volume, 1961 (6th printing, 1980)
(5)	L. Van Hove, N. M. Hugenholtz, and L. P. Howland	Problems in Quantum Theory of Many-Particle Systems: A Lecture Note and Reprint Volume, 1961
(6)	D. Pines	The Many-Body Problem: A Lecture Note and Reprint Volume, 1961 (6th printing, 1982)
(7)	H. Frauenfelder	The Mössbauer Effect: A Review—with a Collection of Reprints, 1962
(8)	L. P. Kadanoff and G. Baym	Quantum Statistical Mechanics: Green's Function Methods in Equilibrium and Nonequilibrium Problems, 1962 (5th printing, 1981)
(9)	G. E. Pake	Paramagnetic Resonance: An Introductory Monograph, 1962 [cr. (42)—2nd edition]
(10)	P. W. Anderson	Concepts in Solids: Lectures on the Theory of Solids, 1963 (5th printing, 1982)
(11)	S. C. Frautschi	Regge Poles and *S*-Matrix Theory, 1963
(12)	R. Hofstadter	Electron Scattering and Nuclear and Nucleon Structure: A Collection of Reprints with an Introduction, 1963
(13)	A. M. Lane	Nuclear Theory: Pairing Force Correlations to Collective Motion, 1964
(14)	R. Omnès and M. Froissart	Mandelstam Theory and Regge Poles: An Introduction for Experimentalists, 1963
(15)	E. J. Squires	Complex Angular Momenta and Particle Physics: A Lecture Note and Reprint Volume, 1963
(16)	H. L. Frisch and J. L. Lebowitz	The Equilibrium Theory of Classical Fluids: A Lecture Note and Reprint Volume, 1964
(17)	M. Gell-Mann and Y. Ne'eman	The Eightfold Way: (A Review—with a Collection of Reprints), 1964
(18)	M. Jacob and G. F. Chew	Strong-Interaction Physics: A Lecture Note Volume, 1964
(19)	P. Nozières	Theory of Interacting Fermi Systems, 1964

FRONTIERS IN PHYSICS

David Pines, Editor *(continued)*

(20)	J. R. Schrieffer	Theory of Superconductivity, 1964 (revised 3rd printing, 1983)
(21)	N. Bloembergen	Nonlinear Optics: A Lecture Note and Reprint Volume, 1965 (4th printing, 1981)
(22)	R. Brout	Phase Transitions, 1965
(23)	I. M. Khalatnikov	An Introduction to the Theory of Superfluidity, 1965
(24)	P. G. deGennes	Superconductivity of Metals and Alloys, 1966
(25)	W. A. Harrison	Pseudopotentials in the Theory of Metals, 1966 (2nd printing, 1971)
(26)	V. Barger and D. Cline	Phenomenological Theories of High Energy Scattering: An Experimental Evaluation, 1967
(27)	P. Choquard	The Anharmonic Crystal, 1967
(28)	T. Loucks	Augmented Plane Wave Method: A Guide to Performing Electronic Structure Calculations—A Lecture Note and Reprint Volume, 1967
(29)	Y. Ne'eman	Algebraic Theory of Particle Physics: Hadron Dynamics in Terms of Unitary Spin Currents, 1967
(30)	S. L. Adler and R. F. Dashen	Current Algebras and Applications to Particle Physics, 1968
(31)	A. B. Migdal	Nuclear Theory: The Quasiparticle Method, 1968
(32)	J. J. J. Kokkedee	The Quark Model, 1969
(33)	A. B. Migdal and V. Krainov	Approximation Methods in Quantum Mechanics, 1969
(34)	R. Z. Sagdeev and A. A. Galeev	Nonlinear Plasma Theory, 1969
(35)	J. Schwinger	Quantum Kinematics and Dynamics, 1970
(36)	R. P. Feynman	Statistical Mechanics: A Set of Lectures, 1972 (7th printing, 1982)
(37)	R. P. Feynman	Photo-Hadron Interactions, 1972
(38)	E. R. Caianiello	Combinatorics and Renormalization in Quantum Field Theory, 1973
(39)	G. B. Field, H. Arp, and J. N. Bahcall	The Redshift Controversy, 1973 (2nd printing, 1976)
(40)	D. Horn and F. Zachariasen	Hadron Physics at Very High Energies, 1973
(41)	S. Ichimaru	Basic Principles of Plasma Physics: A Statistical Approach, 1973 (2nd printing, with revisions, 1980)
(42)	G. E. Pake and T. L. Estle	The Physical Principles of Electron Paramagnetic Resonance, 2nd Edition, completely revised, enlarged, and reset, 1973 [cf. (9)—1st edition]

FRONTIERS IN PHYSICS

David Pines, Editor (*continued*)

Volumes published from 1974 onward are being numbered as an integral part of the bibliography:

Number		
43	R. C. Davidson	Theory of Nonneutral Plasmas, 1974
44	S. Doniach and E. H. Sondheimer	Green's Functions for Solid State Physicists, 1974 (2nd printing, 1978)
45	P. H. Frampton	Dual Resonance Models, 1974 ✓
46	S. K. Ma	Modern Theory of Critical Phenomena, 1976 (5th printing, 1982)
47	D. Forster	Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions, 1975 (3rd printing, 1983)
48	A. B. Migdal	Qualitative Methods in Quantum Theory, 1977
49	S. W. Lovesey	Condensed Matter Physics: Dynamic Correlations, 1980
50	L. D. Faddeev and A. A. Slavnov	Gauge Fields: Introduction to Quantum Theory, 1980 ✓
51	P. Ramond	Field Theory: A Modern Primer, 1981 (4th printing, 1983) ✓
52	R. A. Broglia and A. Winther	Heavy Ion Reactions: Lecture Notes Vol. I: Elastic and Inelastic Reactions, 1981
53	R. A. Broglia and A. Winther	Heavy Ion Reactions: Lecture Notes Vol. II, *in preparation*
54	Howard Georgi	Lie Algebras in Particle Physics: From Isospin to Unified Theories, 1982 ✓
55	P. W. Anderson	Basic Notions of Condensed Matter Physics, 1983
56	Chris Quigg	Gauge Theories of the Strong, Weak, and Electromagnetic Interactions, 1983 ✓
57	S. I. Pekar	Crystal Optics and Additional Light Waves, 1983
58	S. J. Gates, M. T. Grisaru, M. Roček, and W. Siegel	Superspace *or* One Thousand and One Lessons in Supersymmetry, 1983
59	R. N. Cahn	Semi-Simple Lie Algebras and Their Representations, 1984
60	G. G. Ross	Grand Unified Theories, 1984 ✓
61	S. W. Lovesey	Condensed Matter Physics: Dynamic Correlations, 1986
62	P. H. Frampton	Gauge Field Theories, 1986 ✓
63	J. I. Katz	High Energy Astrophysics, 1986
64	T. J. Ferbel	Experimental Techniques in High Energy Physics, 1987
65	T. Appelquist, A. Chodos, and P. G. O. Freund	Modern Kaluza-Klein Theories, 1987 ✓

Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/modernkaluzaklei0000unse

EDITOR'S FOREWORD

The problem of communicating in a coherent fashion recent developments in the most exciting and active fields of physics continues to be with us. The enormous growth in the number of physicists has tended to make the familiar channels of communication considerably less effective. It has become increasingly difficult for experts in a given field to keep up with the current literature; the novice can only be confused. What is needed is both a consistent account of a field and the presentation of a definite "point of view" concerning it. Formal monographs cannot meet such a need in a rapidly developing field, while the review articles rarely receive the priority they deserve in the physics community. Indeed, it would seem that the people most actively engaged in developing a given field are the people least likely to write at length about it.

FRONTIERS IN PHYSICS was conceived in 1961 in an effort to improve the situation in several ways. Leading physicists frequently give a series of lectures, a graduate seminar, or a graduate course in their special fields of interest. Such lectures serve to summarize the present status of a rapidly developing field and may well constitute the only coherent account available at the time. Often, notes on lectures exist (prepared by the lecturer himself, by graduate students, or by postdoctoral fellows) and are distributed in mimeographed form on a limited basis. One of the principal purposes of the FRONTIERS IN PHYSICS Series is to make such notes available to a wider audience of physicists.

The publication of collections of reprints of recent articles in very active fields of physics also improves communication between physicists. The value of the reprints is, however, enhanced when the collection is accompanied by an introduction of moderate length which ties the collection together, and in so doing constitutes a brief survey of the present status of the field. In the present volume, the editors, Thomas W. Appelquist, Alan Chodos, and Peter G. O. Freund, have accomplished this task in admirable fashion for the field of higher dimensional unification, a field to which they themselves have made significant contributions.

Modern Kaluza-Klein Theories will be of interest to everyone concerned with understanding the attempts which theoretical physicists have made during the present century to construct a unified theory of all known interactions between elementary particles. The authors place Kaluza-Klein theory, a seminal attempt to solve the unification problem by utilizing a higher-dimensional setting, in its proper historical context. Thus their collection begins with papers of Nordström, Kaluza, and Klein, written over sixty years ago, and concludes with the brief account, written in 1984, by Michael Green and John Schwartz of their exciting work on superstrings in ten-dimensional space time. It contains many of the key papers on supergravity and Kaluza-Klein cosmologies, and should constitute a valuable reference volume for all theorists engaged in, or contemplating, research on superstrings.

David Pines
Urbana, Illinois

September, 1986

PREFACE

With electroweak theory and chromodynamics freshly constructed, new momentum was imparted to the age old endeavour of unifying all known interactions and forms of matter. This had to include gravity and fermions as well. Gauged supersymmetry in 4 dimensions was an obvious tool, but it produced rather complicated "extended" theories, which met with little phenomenological success. This has led to a reconsideration of the old (1920s) idea of unification in a higher dimensional setting, with the extra space-dimensions dynamically compactified. There has been an intense exploration of these theories both with theoretical aims (finding a unique unification, obtaining a consistent quantum theory of gravity) and phenomenological aims (predicting the grand unification group and the low energy particle spectrum, as well as finding cosmological implications). This work seems now to have arrived at a most exciting stage involving extended objects, super-strings, in a ten-dimensional space-time.

In this volume we have collected a set of important and, we believe, representative papers on higher dimensional unification. We have also included our own brief survey of the field with extensive references to the existing literature. Obviously space limitations have prevented us from including some further interesting papers. Still, we hope this collection will be of use for those about to embark on research in this exciting field.

Thomas W. Appelquist
Alan Chodos
Peter G. O. Freund

July, 1986

TABLE OF CONTENTS

SERIES LISTING v

EDITOR'S FOREWORD ix

PREFACE xi

CHAPTER I. INTRODUCTION 1

CHAPTER II. HISTORY 49

GUNNAR NORDSTRÖM, "ON THE POSSIBILITY OF A UNIFICATION OF THE ELECTROMAGNETIC AND GRAVITATION FIELDS" 50

GUNNAR NORDSTRÖM, "ÜBER DIE MÖGLICHKEIT, DAS ELEKTROMAGNETISCHE FELD UND DAS GRAVITATIONSFELD ZU VEREINIGEN" 57

TH. KALUZA, "ON THE UNITY PROBLEM OF PHYSICS" 61

TH. KALUZA, "ZUM UNITÄTSPROBLEM DER PHYSIK" 69

O. KLEIN, "QUANTUM THEORY AND FIVE DIMENSIONAL THEORY OF RELATIVITY" 76

O. KLEIN, "THE ATOMICITY OF ELECTRICITY AS A QUANTUM THEORY LAW" 88

A. EINSTEIN AND P. BERGMANN, "ON A GENERALIZATION OF KALUZA'S THEORY OF ELECTRICITY" 89

Y. THIRY, "THE EQUATIONS OF KALUZA'S UNIFIED THEORY" 108

Y. THIRY, "LES ÉQUATIONS DE LA THÉORIE UNITAIRE DE KALUZA" 110

CHAPTER III. NON-ABELIAN GENERALIZATION 113

B. S. DEWITT, "DYNAMICAL THEORY OF GROUPS AND FIELDS", PROBLEM NUMBER 77, 114

RYSZARD KERNER, “GENERALIZATION OF THE KALUZA-KLEIN THEORY FOR AN ARBITRARY NON-ABELIAN GAUGE GROUP” 115

Y. M. CHO AND P. G. O. FREUND, “NON-ABELIAN GAUGE FIELDS AS NAMBU-GOLDSTONE FIELDS” 125

E. CREMMER AND J. SCHERK, “SPONTANEOUS COMPACTIFICATION OF SPACE IN AN EINSTEIN-YANG-MILLS-HIGGS MODEL” 135

J. F. LUCIANI, “SPACE-TIME GEOMETRY AND SYMMETRY BREAKING” 143

ABDUS SALAM AND J. STRATHDEE, “ON KALUZA-KLEIN THEORY” 163

CHAPTER IV. SUPERGRAVITY 201

E. CREMMER, B. JULIA AND J. SCHERK, “SUPERGRAVITY THEORY IN 11 DIMENSIONS” 202

E. CREMMER AND B. JULIA, “THE $N=8$ SUPERGRAVITY THEORY. I. THE LAGRANGIAN” 206

P. G. O. FREUND AND MARK A. RUBIN, “DYNAMICS OF DIMENSIONAL REDUCTION” 210

F. ENGLERT, “SPONTANEOUS COMPACTIFICATION OF ELEVEN-DIMENSIONAL SUPERGRAVITY” 213

I. BARS AND S. W. MACDOWELL, “GRAVITY WITH EXTRA GAUGE SYMMETRY” 217

B. BIRAN, F. ENGLERT, B. DE WIT AND H. NICOLAI, “GAUGED $N=8$ SUPERGRAVITY AND ITS BREAKING FROM SPONTANEOUS COMPACTIFICATION” 220

A. CASHER, F. ENGLERT, H. NICOLAI AND M. ROOMAN, “THE MASS SPECTRUM OF SUPERGRAVITY ON THE ROUND SEVEN-SPHERE” 226

M. A. AWADA, M. J. DUFF AND C. N. POPE, “$N=8$ SUPERGRAVITY BREAKS DOWN TO $N=1$” 242

M. J. DUFF, B. E. W. NILSSON AND C. N. POPE, “SPONTANEOUS SUPERSYMMETRY BREAKING BY THE SQUASHED SEVEN-SPHERE” 246

R. D'AURIA, P. FRE AND P. VAN NIEUWENHUIZEN, "SYMMETRY BREAKING IN $d=11$ SUPERGRAVITY ON THE PARALLELIZED SEVEN-SPHERE" 251

FEZA GÜRSEY AND CHIA-HSIUNG TZE, "OCTONIONIC TORSION ON S^7 AND ENGLERT'S COMPACTIFICATION OF $d=11$ SUPERGRAVITY" 258

G. F. CHAPLINE AND N. S. MANTON, "UNIFICATION OF YANG-MILLS THEORY AND SUPERGRAVITY IN TEN DIMENSIONS" 264

D. Z. FREEDMAN, G. W. GIBBONS AND P. C. WEST, "TEN INTO FOUR WON'T GO" 269

B. DE WIT AND H. NICOLAI, "A NEW SO(7) INVARIANT SOLUTION OF $d=11$ SUPERGRAVITY" 271

CHAPTER V. TOWARDS A REALISTIC KALUZA-KLEIN THEORY 277

EDWARD WITTEN, "SEARCH FOR A REALISTIC KALUZA-KLEIN THEORY" 278

CHAPTER VI. COSMOLOGY 295

ALAN CHODOS AND STEVEN DETWEILER, "WHERE HAS THE FIFTH DIMENSION GONE?" 296

PETER G. O. FREUND, "KALUZA-KLEIN COSMOLOGIES" 300

ENRIQUE ALVAREZ AND M. BELEN GAVELA, "ENTROPY FROM EXTRA DIMENSIONS" 311

Q. SHAFI AND C. WETTERICH, "COSMOLOGY FROM HIGHER-DIMENSIONAL GRAVITY" 315

DESHDEEP SAHDEV, "TOWARDS A RELATIVISTIC KALUZA-KLEIN COSMOLOGY" 320

EDWARD W. KOLB AND RICHARD SLANSKY, "DIMENSIONAL REDUCTION IN THE EARLY UNIVERSE: WHERE HAVE THE MASSIVE PARTICLES GONE?" 325

KEI-ICHI MAEDA, "IS COSMOLOGICAL DIMENSIONAL REDUCTION POSSIBLE?" 330

S. RANDJBAR-DAEMI, ABDUS SALAM AND J. STRATHDEE, "ON KALUZA-KLEIN COSMOLOGY" 335

CHAPTER VII. QUANTUM EFFECTS 341

H. B. G. CASIMIR, "ON THE ATTRACTION BETWEEN TWO PERFECTLY CONDUCTING PLATES" 342

THOMAS APPELQUIST AND ALAN CHODOS, "QUANTUM DYNAMICS OF KALUZA-KLEIN THEORIES" 345

STEVEN WEINBERG, "CHARGES FROM EXTRA DIMENSIONS" 359

MARK A. RUBIN AND BERNARD D. BARKANROTH, "TEMPERATURE EFFECTS IN FIVE-DIMENSIONAL KALUZA-KLEIN THEORY" 364

PHILIP CANDELAS AND STEVEN WEINBERG, "CALCULATION OF GAUGE COUPLINGS AND COMPACT CIRCUMFERENCES FROM SELF-CONSISTENT DIMENSIONAL REDUCTION" 375

N. A. VORONOV AND YA I. KOGAN, "SPONTANEOUS COMPACTION IN KALUZA-KLEIN MODELS AND THE CASIMIR EFFECT" 420

G. W. GIBBONS AND H. NICOLAI, "ONE-LOOP EFFECTS ON THE ROUND SEVEN-SPHERE" 424

T. INAMI AND K. YAMAGISHI, "VANISHING QUANTUM VACUUM ENERGY IN ELEVEN-DIMENSIONAL SUPERGRAVITY ON THE ROUND SEVEN-SPHERE" 431

CHAPTER VIII. THE PROBLEM OF CHIRAL FERMIONS 437

EDWARD WITTEN, "FERMION QUANTUM NUMBERS IN KALUZA-KLEIN THEORY" 438

C. WETTERICH, "DIMENSIONAL REDUCTION OF FERMIONS IN GENERALIZED GRAVITY" 512

S. RANDJBAR-DAEMI, ABDUS SALAM AND J. STRATHDEE, "SPONTANEOUS COMPACTIFICATION IN SIX-DIMENSIONAL EINSTEIN-MAXWELL THEORY" 542

STEVEN WEINBERG, "QUASI-RIEMANNIAN THEORIES OF GRAVITATION IN MORE THAN FOUR DIMENSIONS" 564

M. GELL-MANN AND B. ZWIEBACH, "CURLING UP TWO SPATIAL DIMENSIONS WITH SU(1,1)/U(1)" 569

CHAPTER IX. MONOPOLES **573**
RAFAEL D. SORKIN, "KALUZA-KLEIN MONOPOLE" **574**

DAVID J. GROSS AND MALCOLM J. PERRY, "MAGNETIC MONOPOLES IN KALUZA-KLEIN THEORIES" **578**

MALCOLM J. PERRY, "NON-ABELIAN KALUZA-KLEIN MONOPOLES" **598**

CHAPTER X. SUPERSTRINGS **603**
M. B. GREEN AND J. H. SCHWARZ, "SUPERSYMMETRICAL STRING THEORIES" **604**

M. B. GREEN AND J. H. SCHWARZ, "ANOMALY CANCELLATIONS IN SUPERSYMMETRIC $d=10$ GAUGE THEORY AND SUPERSTRING THEORY" **608**

CREDITS **615**

CHAPTER

1

Introduction

1.1 The Idea of Higher-Dimensional Unification

The quest for unity in physics is an old one. Already in the last century two basic forces of nature, electricity and magnetism, were unified starting with the work of Ampère and Faraday and culminating in Maxwell's electrodynamics. Once the special relativistic invariance of Maxwell's theory was recognized, it became clear through Minkowski's work that the unification of electricity and magnetism entails a unification of three-dimensional space and of time into a four-dimensional "space-time continuum". There is a parallelism between the electricity-magnetism unification on the one hand, and the space-time unification on the other hand. Just as the space and time coordinates now jointly describe a higher four-dimensional manifold, so three-dimensional quantities like the electric and magnetic fields $\vec{E}$ and $\vec{H}$ become components of a single six-component antisymmetric tensor $F_{\mu\nu}$, and the corresponding potentials ϕ and $\vec{A}$ become unified into a four-vector.

As soon as a relativistic theory of gravitation became available, the question of its unification with Maxwell's theory arose. Even before Einstein's general relativity the Finnish physicist Gunnar Nordström[1] proposed a relativistic theory in which gravity was described by a scalar field coupled to the trace of the energy momentum tensor. In 1914, still before general relativity, Nordström[2*] then proceeded to unify his theory of gravitation with Maxwell's theory in a most imaginative way. Inspired by Minkowski's four-dimensional space-time continuum, Nordström added yet another space dimension, thus obtaining a flat five-dimensional world. There he introduced an abelian five-vector gauge field for which he wrote down the Maxwell equations including a conserved five-current. He then identified the fifth component of the five-vector potential with scalar gravity, whereas the first four components of the five-vector potential he identified with the Maxwell four-potential. With these interpretations he then noticed that in the cylindrical case (when all dynamical variables become independent of the fifth coordinate) the equations of his five-dimensional "Maxwell"-theory reduced to those of the four-dimensional Maxwell-Nordström electromagnetic-gravitational theory. It is then fair to say that higher-dimensional unification starts with Nordström, who assumed scalar gravity in our four-dimensional world to be a remnant of an abelian gauge theory in a five-dimensional flat space-time. Remarkably, the most recent work on superstrings[3*] incorporates both the ideas of Nordström and the subsequent ideas of Kaluza as we shall see.

The next step was taken by the mathematician Theodor Kaluza in 1919 in the wake of Einstein's general relativity. Kaluza[4*] proposed that one pass to an Einstein-type theory of gravity in five dimensions, from which ordinary four-dimensional Einstein gravity and

Maxwell electromagnetism are to be obtained upon imposing a cylindrical constraint.

More precisely what this amounts to is to start with a five-dimensional manifold **M** which is the product **M** = $M_4 x S^1$ of a four-dimensional space-time M_4 with a circle S^1. The metric $\gamma_{MN}(x,y)$ on the five-manifold **M** (M,N = 0,1,2,3,5) is a function of both the coordinates x^μ (μ=0,1,2,3) on M_4 and of $y \equiv x^5$ the coordinate on the circle S^1 (throughout this section capital letter indices run over the values 0,1,2,3,5 and Greek indices over the values 0,1,2,3). It is convenient to replace the fifteen field variables $\gamma_{MN}(=\gamma_{NM})$ by fifteen new field variables $g_{\mu\nu}=g_{\nu\mu}, A_\mu, \phi$ according to the field redefinitions

$$\begin{aligned} \gamma_{\mu\nu} &= g_{\mu\nu} + e^2\kappa^2\phi A_\mu A_\nu \\ \gamma_{\mu 5} &= \gamma_{5\mu} = e\kappa\phi A_\mu \qquad (1.1) \\ \gamma_{55} &= \phi \end{aligned}$$

All field quantities, old and new, are periodic functions of the coordinate y on the circle. If $y=\rho\theta$, where θ is the usual angular coordinate and ρ the radius of the circle, then the period is $2\pi\rho$. Thus any field quantity F(x,y) (F being any of the $g_{\mu\nu}$'s, A_μ's, ϕ, or γ_{MN}'s) admits a Fourier expansion

$$F(x,y) = \sum_{n=-\infty}^{+\infty} F^{(n)}(x)e^{iny/\rho} \qquad (1.2)$$

Kaluza assumed the five-dimensional dynamics to be governed by a gravitational Einstein-Hilbert action

$$I_5 = -\frac{1}{16\pi G_5}\int \sqrt{|\gamma_5|}\ R_5 d^5x, \qquad (1.3)$$

with $\gamma_5=\det(\gamma_{MN})$, R_5 the five-dimensional curvature scalar and G_5 a five-dimensional counterpart of the gravitational constant. Using the Fourier expansions, the y-dependence becomes explicit so that the y-integration can be carried out. A four-dimensional action involving an infinity of fields - the Fourier components $A_{\mu}^{(n)}$, $g_{\mu\nu}^{(n)}(x)$, $\phi^{(n)}$-then emerges. At this point Kaluza imposed* a "cylindricity" condition: he truncated the action by dropping all harmonics with n≠0, retaining only the zero modes:

$$g_{\mu\nu}(x,y)=g_{\mu\nu}^{(0)}(x);\ A_{\mu}(x,y)=A_{\mu}^{(0)}(x);\ \phi(x,y)=\phi^{(0)}(x). \quad (1.4)$$

The five-dimensional line element then takes the form

$$ds_5^2 \equiv \gamma_{MN}dx^M dx^N = ds_4^2 + \phi^{(0)}(x)(dx^5 + e\kappa A_{\mu}^{(0)}(x)dx^{\mu})^2\ , \quad (1.5a)$$

where

$$ds_4^2 \equiv g_{\mu\nu}^{(0)}(x)dx^{\mu}dx^{\nu} \quad (1.5b)$$

is the four-dimensional line element corresponding to the metric $g_{\mu\nu}^{(0)}(x)$. The line element (1.5) is invariant under the transformations

$$\begin{aligned} x_5^{\mu} &\rightarrow x_5^{\mu} \\ x &\rightarrow x + e\kappa\alpha(x^{\rho}) \\ A_{\mu}^{(0)} &\rightarrow A_{\mu}^{(0)} - \partial_{\mu}\alpha(x^{\rho}) \\ \phi^{(0)} &\rightarrow \phi^{(0)} \\ g_{\mu\nu}^{(0)} &\rightarrow g_{\mu\nu}^{(0)}, \end{aligned} \quad (1.6)$$

---

* Historically, Kaluza proceeded to this step without all the harmonic expansion. He even made the additional assumption $\phi^{(0)}(x) = 1$.

which we recognize as Abelian gauge transformation a la Weyl. Here these transformations assume a geometrical meaning as shifts in the fifth coordinate by an amount $\alpha(x^\rho)$, which depends on ordinary four-space-time. The Abelian gauge symmetry in four dimensions originates in the isometries of the small circle in the fifth dimension.

When the **y** integration is carried out with the cylindric truncation enforced, the action (1.3), invariant under five-dimensional general coordinate transformation, reduces to a four-dimensional action invariant under both four-dimensional general coordinate transformations and abelian gauge transformations. This four-dimensional action is, up to a surface term,

$$I_4 = \int \sqrt{|g_4^{(0)}|}\,\sqrt{|\phi^{(0)}|}\,\Big(-\frac{1}{16\pi G}\,R_4^{(0)} + \frac{e^2\kappa^2}{16\pi G}\,\phi^{(0)} g^{(0)\mu\rho} g^{(0)\nu\sigma} F^{(0)}_{\mu\nu} F^{(0)}_{\rho\sigma}\Big) \qquad (1.7a)$$

with

$$G = \frac{G_5}{2\pi\rho},\; g_4^{(0)} = \det(g^{(0)}_{\mu\nu}),\; F^{(0)}_{\mu\nu} = \partial_\mu A^{(0)}_\nu - \partial_\nu A^{(0)}_\mu, \text{ and}$$

$$R_4^{(0)} = \text{scalar curvature calculated from the four-metric } g^{(0)}_{\mu\nu} \qquad (1.7b)$$

(our metric convention calls for a minus (plus) sign for a time (space)-like dimension).

This action involves a graviton ($g^{(0)}_{\mu\nu}$), an Abelian gauge boson ($A^{(0)}_\mu$) and a scalar field $\phi^{(0)}$. Kaluza arbitrarily set

$$\phi^{(0)} = \text{constant}, \qquad (1.8)$$

in which case I_4 turns into the four-dimensional Einstein-

Maxwell action. To be sure, one has to have $\phi^{(0)}>0$ in order to have the proper relative sign of the Einstein and Maxwell terms, so that energy is positive. This in turn means that the fifth dimension must be space-like. As we shall see below this is a general feature: the extra dimensions must all be space-like. This matches one's expectations based on causality (more than one time-like dimension would lead to closed time-like curves).

Imposing the Kaluza constraint (1.8) leads to difficulties. For the scalar field equation together with the trace of the Einstein equation then imply the unwanted constraint

$$F^{(0)}_{\mu\nu}\, F^{(0)\mu\nu} = 0 \tag{1.9}$$

in the Maxwell field. Kaluza avoided this, imposing (1.8) directly in the action (1.7) and never taking the variation with respect to $\phi^{(0)}$, a somewhat "ostrich-like" strategy. The necessity of including the scalar field was not pointed out until the nineteen-forties when Jordan[5,6] and later independently Thiry[7*] considered the full action (1.7).

In addition to the invariances under general coordinate transformations and gauge transformations, the action (1.7) also exhibits an invariance under global scale transformations:

$$\begin{aligned} g^{(0)}_{\mu\nu} &\to \lambda^{-1} g^{(0)}_{\mu\nu} \\ A^{(0)}_{\mu} &\to \lambda^{-3/2} A^{(0)}_{\mu} \\ \phi^{(0)} &\to \lambda^{2} \phi^{(0)} \quad . \end{aligned} \tag{1.10}$$

The field equations of the original five-dimensional theory have a solution in which the five-dimensional space-time is the direct product of a circle with flat four-dimensional

Minkowski space-time. Then

$$g_{\mu\nu} = \eta_{\mu\nu},\ A_\mu = 0,\ \phi = 1 \qquad (1.11)$$

($\eta_{\mu\nu}$ is the four-dimensional Minkowski metric (-1,+1,+1,+1)). This solution serves as a natural vacuum, and it spontaneously breaks the scale invariance (1.10). The massless $\phi^{(0)}$-field is the Nambu-Goldstone boson associated with this spontaneous symmetry breaking. So the zero mode spectrum includes spin 2 and spin 1 gauge fields and a spin 0 Nambu-Goldstone boson. In the full quantum theory the spin zero boson is expected to acquire a mass.

Of course, the full classical theory contains not only the zero modes, but also the n≠0 harmonics (Eq. (1.2)). The action (1.5) determines their spins, masses and couplings. They all have spins ≤ 2, and they all are massive. The nth harmonics have mass

$$m_n = |n|/\rho, \qquad (1.12)$$

where ρ, as before, is the radius of the small circle in the fifth dimension. The couplings of these harmonics to the gauge field $A_\mu^{(0)}$ are determined from the action (1.3), and these harmonics do carry the "electric" charge. The nth harmonic carries electric charge

$$g_n = n\,\frac{4\sqrt{\pi G}}{\rho}\ . \qquad (1.13)$$

Remarkably, electric charge is quantized[8*], because the fifth dimension is compact. We see that the elementary charge is

$$e = \frac{4\sqrt{\pi G}}{\rho} \qquad (1.14a)$$

and the corresponding fine structure constant is

$$\alpha = \frac{4G}{\rho^2} \, . \tag{1.14b}$$

If α is to correspond to the U(1) subgroup of a grand-unification group then $\alpha \sim \frac{1}{100}$ so that the circumference of the small circle $\ell \equiv 2\pi\rho \sim 100 \sqrt{G} \sim 10^{-17}$ GeV^{-1}. The circle must be very small indeed; a size about 100 Planck lengths could hardly have been detected as yet. Nevertheless, this is large enough to call into question grand-unification in four dimensions: the scales at which the grand unification group is to reveal itself unbroken are close to the scales at which the extra dimensions would become manifest.

To make all this applicable in a world with strong and electroweak interactions, of course, one has to introduce more than one extra dimension, and we shall see in Section 3 how this is to be done. First we recount some of the colorful history of Kaluza-Klein theories.

1.2 **Early History**

Although Nordström who first suggested going to five dimensions[2*] was a distinguished physicist in his own right, his work, published in a widely read physics journal, was not referred to by Kaluza or by any of the many later authors. In fact, it first resurfaced in a recent historical work.[9] That Kaluza should not have known of Nordström's work is quite plausible: he was a mathematician, after all. That Einstein, familiar with the rest of Nordström's work, should have been unaware of the latter's foray into five dimensions is surprising. It is known that Einstein's opinion of Nordström's theory of gravitation was not particularly high[9] and this may explain his disregard of any work in which Nordström gravity figures prominently. There are also reasons to believe that the personal relationship of the two men was less than cordial.[10] Kaluza himself submitted his work to Einstein in 1919, who after repeated vacillation[9] as to its merit finally decided to sponsor its publication in 1921.

In 1926 Oskar Klein[11*] and the Russian physicist H. Mandel[12] independently rediscovered Kaluza's theory. Klein noted the quantization of the electric charge[8*] and hoped Kaluza theory would underlie quantum mechanics. The relativistic generalization of Schrödinger's equation was carried out[13] independently by many authors: Schrödinger[14], Klein[11], Gordon[15], Fock[16], de Donder and van den Dungen[17], Kudar[18]. This equation, now commonly known as the Klein-Gordon equation, was arrived at by both Klein and Fock starting from Kaluza theory: a zero mass wave equation in five dimensions yields four-dimensional Klein-Gordon equations for the individual harmonics. It is amusing that the Mandel and Fock papers submitted only weeks apart indicate their authors as residing in St. Petersburg and Leningrad respectively.

All this early work is rather apologetic in tone; the fifth dimension is viewed as a mathematical trick devoid of any physical significance. Oskar Klein[8*] comes closest to the modern point of view: he discusses the higher harmonics and the size of the small circle. Later Einstein and Bergmann[19*] also adopted such a point of view.

To emphasize the purely mathematical nature of the idea, Hoffmann and Veblen[20] proposed a projective interpretation of the fifth coordinate. This approach was developed by Pauli[21], Pais[22], Jordan[5], Schmutzer[23], Ludwig[24] and others.

It is amazing that in all this early work and in particular in the three independent discoveries of the five-dimensional theory by Kaluza, Klein and Mandel the scalar field $\phi^{(0)}$ is always set equal to one, without much worry about the constraint (1.10). Jordan[5] appears to have been the first to realize the importance of including the scalar field $\phi^{(0)}$. His original paper on the subject did not appear in print, as the journal Physikalische Zeitschrift (the same journal that carried Nordström's original 1914 paper) was bombed out of existence shortly before Jordan was to return the galley proofs. Somehow a copy of these galley proofs made it to Pauli in Princeton, who circulated them to Einstein and Bergmann, themselves involved in similar work. There is a nice account by Bergmann[6] of the parallel Jordan and Bergmann-Einstein work on the inclusion of the scalar field $\phi^{(0)}$. Again, years later, Thiry[7*], in a very clear and elegant paper, noticed independently the effect of including the scalar field. Jordan himself developed[25] a whole new theory of gravity and cosmology in which gravity involves both spin 2 and spin 0 fields. Difficulties in this theory were noted by Fierz[26], but the main idea was further developed by Brans and Dicke.[27]

The fifth dimension has continued to hold a fascination for physicists. Einstein returned to this idea over and over.[9] A number of other people worked on it: Rumer[28], Souriau[29], Rayski[30]. Most remarkable is Klein's derivation[31] (from the five-dimensional theory!) of a vector field theory very similar to that of Yang and Mills.

1.3 Beyond Five Dimensions

There was no real reason to extend the Kaluza-Klein idea beyond five dimensions until the emergence of non-Abelian gauge field theories. From the time of their invention by Yang and Mills in 1954 until their great blossoming in the early 1970's, these theories were studied intermittently by only a few intrepid pioneers. In the middle of this era, in 1963, DeWitt[32*] suggested that a unification of Yang-

Mills theories and gravitation could be achieved in a higher dimensional Kaluza-Klein framework. The idea appears as an assigned problem in his 1963 Les Houches lectures. Trautman[33] was independently aware of this possibility, as were others.

A detailed discussion of the Kaluza-Klein unification of gravity and Yang-Mills theories, including the correct form of the 4+N dimensional metric, first appears in the work of Kerner.[34*] The first complete derivation of the four-dimensional gravitational plus Yang-Mills plus scalar theory from a (4+N)-dimensional Einstein-Hilbert action was finally given by Cho and Freund[35*,36] in 1975.

The weakness of this higher-dimensional work was the absence of any good reason as to **why** any dimensions would compactify, let alone the right number, so as to leave the ordinary four-dimensional "large" world. While the five-dimensional theory at least admitted the compactified fifth dimension along with Minkowski space as a solution to the five-dimensional equations of motion, even this was not true of the higher-dimensional theories.

The essential reason for this is that the higher dimensional manifolds that give rise to Yang-Mills theories have curvature. If a 4+N dimensional Einstein theory is to compactify into the direct product of four-dimensional space-time M_4 and a compact internal space with isometries, the metric $\gamma_{MN}(x,y)$ can be written as follows in the zero-mode approximation (in order to make the notation more compact, in this section we do not distinguish zero-mode quantities by a superscript (0), as was done in the previous section):

$$\gamma_{MN}(x,y) =$$

$$\left[\begin{array}{c|c} g_{\mu\nu}(x) + \gamma_{mn}(y)\zeta^m_\alpha(y)\zeta^n_\beta(y)A^\alpha_\mu(x)A^\beta_\nu(x) & \gamma_{mn}(y)\zeta^m_\alpha(y)A^\alpha_\mu(x) \\ \hline \gamma_{mn}(y)\zeta^n_\beta(y)A^\beta_\nu(x) & \gamma_{mn}(y) \end{array}\right] \tag{3.1}$$

The metric $\gamma_{mn}(y)$ is that of the corresponding N-dimensional symmetric space and the Killing vectors $\zeta^n_\alpha(y)$ have upper indices running over the N dimensions and lower indices running over the dimension of the symmetry group. They obey the relation

$$\zeta^m_\beta \, \partial_m \, \zeta^n_\gamma - \zeta^m_\gamma \, \partial_m \zeta_\beta = f^\alpha_{\beta\gamma} \, \zeta^n_\alpha \tag{3.2}$$

where $f^\alpha_{\beta\gamma}$ are the structure constants of the group.

To see the difficulty of a realistic compactification, we insert the above metric into the 4+N dimensional Einstein action

$$I_{4+N} = - \frac{1}{16\pi \, G_{4+N}} \int \sqrt{|\gamma_{4+N}|} \, (R_{4+N} + \Lambda) \, d^{4+N}x, \tag{3.3}$$

where Λ is a cosmological constant. The resulting four-dimensional Lagrangian is

$$\begin{aligned} L_4 = - \frac{1}{16\pi \, G_{4+N}} \int d^N y \sqrt{|\gamma_4(x)|} \; \sqrt{|\gamma_N(y)|} \; \{R_4(x) \\ + R_N(y) + \Lambda \qquad\qquad (3.4) \\ + \frac{1}{4} \gamma_{mn}(y) \; \zeta^m_\alpha(y) \; \zeta^n_\beta(y) \; F^\alpha_{\mu\nu}(x) \; F^\beta_{\lambda\rho}(x) \; g^{\mu\lambda}(x) g^{\nu\rho}(x)\}, \end{aligned}$$

where $R_4(x)$ and $R_N(y)$ are the four- and N-dimensional scalar curvatures and

$$F^\alpha_{\mu\nu} = \partial_\mu A^\alpha_\nu - \partial_\nu A^\alpha_\mu + f^{\alpha\beta\gamma} \, A^\beta_\mu A^\gamma_\nu \tag{3.5}$$

The four-dimensional Newton's constant is defined by

$$\frac{1}{16\pi G_4} = \frac{1}{16\pi G_{4+N}} \int d^N y \left| \gamma_N(y) \right|^{1/2} \equiv \frac{1}{16\pi G_{4+N}} \Omega_N \ . \qquad (3.6)$$

The last term in L_4 can be recognized as the Yang-Mills Lagrangian once the integration over y is performed. Defining

$$\int d^N y \ \gamma_N^{1/2}(y)\gamma_{mn}(y)\zeta^n_\alpha(y) \ \zeta^m_\beta(y) \ \equiv \frac{16\pi G_4}{g^2} \ \Omega_N \delta\alpha\beta, \qquad (3.7)$$

the Yang-Mills term is obtained, with g^2 the gauge coupling constant.

The problem with this attractive scenario is that four-space cannot be flat. In the "vacuum", with gauge fields set to zero, the Einstein equations are

$$R_{MN} - \frac{1}{2} \ \gamma_{MN} \ (R + \Lambda) = 0, \qquad (3.8)$$

where the curvature R and the Ricci tensor R_{MN} are formed from the metric

$$\gamma_{MN} = \left(\begin{array}{c|c} \gamma_{\mu\nu}(x) & 0 \\ \hline 0 & \gamma_{mn}(y) \end{array} \right) \qquad (3.9)$$

If four-space is to be flat, $R_{MN}=0$ for the space time indices, and therefore $R+\Lambda=0$. But then R_{MN} must vanish for the internal indices as well and this cannot be the case if the internal space is curved.

Cremmer and Scherk[37*] began to address this problem by pointing out that the inclusion of additional Yang-Mills and scalar matter fields in the higher-dimensional theory would allow classical solutions in which space-time is the direct product of Minkowski space and a compact internal space of constant curvature. This "spontaneous compactification" was achieved, however, by going beyond the pure Kaluza-Klein framework and including extra fields in just such a way as to induce the desired compactification. This program of seeking solutions to the combined Einstein-Yang-Mills equations in 4+D dimensions was continued by others[38], in particular Luciani[39*] who generalized the work of Cremmer and Sherk[37*] to a larger class of internal spaces. Luciani also anticipated later work on Kaluza-Klein supergravity by observing that the Einstein-Yang-Mills theory appears as the bosonic sector of an extended supergravity theory. Other approaches to the classical Kaluza-Klein problem[40-43], such as the inclusion of torsion, were explored at about the same time. The possiblity of effecting compactification by the addition of higher derivative terms to the Einstein action was suggested by Wetterich.[44]

A distillation of much of this work on classical, purely bosonic Kaluza-Klein theories was provided by Salam and Strathdee.[45*] With any compact internal space, it is important to classify the four-dimensional fields in terms of a normal-mode expansion appropriate to the symmetry group of the internal space. Salam and Strathdee examined this expansion, including both the massless modes which populate the large-distance four-dimensional world and the massive modes which are not directly observable if the size of the compact manifold is sufficiently small. The role of these modes was further studied by Dolan and Duff[46], who exposed an infinite dimensional Kac-Moody symmetry that can be used to relate them.

All this work on classical, higher-dimensional Kaluza-Klein theories provided a springboard for the study of both Kaluza-Klein supergravity and the quantum dynamics of Kaluza-Klein theories.

1.4 Supergravity in Eleven Dimensions

Ever since Newton, Physics has dealt with two basic concepts: matter and force. As a rule the basic strong, electroweak and gravitational forces have bosonic quanta. On the other hand the basic quanta of matter can be both fermionic and bosonic. In quantum theory the exchange of Fermi matter quanta also produces forces, which unlike basic forces, involve a change of the identity of the source (e.g., by exchanging an electron between an electron and a photon, the original electron turns into a photon and the original photon into an electron), and these forces as a rule do not follow from a simple gauge principle. This distinction between matter and force had long bothered Einstein, since in general relativity the field equations involve the curvature, an eminently geometric object on the force side, and the quite arbitrary energy momentum tensor on the matter-source side. Could matter and force be unified as different components of the same agency? Remarkably, this question receives an affirmative answer in certain supergravity and superstring theories.

Here we deal with the supergravity case. The maximal - and in some sense the master - supergravity is the one in eleven dimensions constructed by Cremmer, Julia and Scherk[47]*. Roughly speaking there are no supergravities beyond eleven-dimensions[48] because the gravitino, a Rarita-Schwinger spin-vector, in twelve or more dimensions has more than 128 degrees of freedom. When reduced to four dimensions on a torus, this would imply a supergravity with $N>8$, containing therefore fields with spin larger that two; this is generally believed to lead to inconsistencies. In eleven dimensions the Majorana gravitino has precisely 128 degrees of freedom, whereas the graviton has only 44. The number of Bose and Fermi degrees of freedom having to be equal in supersymmetric theory, this means that Bose fields beyond gravity have to appear in eleven-dimensional supergravity. In fact supersymmetry dictates that the missing Bose degrees of freedom be supplied in the form of a massless antisymmetric tensor field with three indices A_{MNP} which indeed has $(\frac{11-2}{3}) = 84 = 128 - 44$ degrees of freedom[47*]. Moreover in eleven dimensions there exist no matter and no Yang-Mills supermultiplets, so that besides gravity one only has its supersymmetric partner A_{MNP} and gravitino-fields as "matter". The source of gravity is thus fixed by supersymmetry. Force and matter uniquely determine each other (the tacit assumption here is the existence of only one graviton); they are but different components of the same supermultiplet. In ten dimensions a similar argument can be made, but there we encounter Yang-Mills supermultiplets whose gauge group is fixed, though not uniquely, by the requirement of anomaly cancellation.[3*] For superstring theories similar considerations apply.[3*]

To find the possible vacua of the eleven-dimensional theory we proceed following Freund and Rubin.[49*] We look for a solution of the classical equations in which the eleven-dimensional world manifold M_{11} is of the form $M_{11} = M_d x M_{11-d}$ where M_d is the space-time and M_{11-d} the small compact manifold. In the vacuum we require the space-time M_d to be maximally symmetric. This then fixes the metric in the form[49*,50,51]

$$\gamma_{MN} = \begin{bmatrix} h(y)g_{\mu\nu}(x) & 0 \\ 0 & g_{mn}(y) \end{bmatrix} \qquad (4.1)$$

where $\mu,\nu = 0,1,2,\ldots d-1$, $m,n = d+1,d+2,\ldots,11$, $x^\mu(y^n)$ are the coordinates on $M_d(M_{11-d})$ and $h(y)$ is a scalar function on the y's. The antisymmetric tensor potential A_{MNP} has its own gauge invariance under

$$A_{MNP}(x,y) \to A_{MNP}(x,y)+\partial_M \alpha_{NP}(x,y)+\partial_N \alpha_{PM}(x,y) + \partial_P \alpha_{MN}(x,y) \qquad (4.2)$$

with $\alpha_{MN} = -\alpha_{NM}$. The corresponding gauge invariant quantities are the field strengths F_{MNPR} given by the curl of A_{MNP}

$$F_{MNPR} = \partial_M A_{NPR} + \partial_N A_{PRM} + \partial_P A_{RMN} + \partial_R A_{MNP} \ . \qquad (4.3)$$

If F or its dual $^*F$ is to have a nonvanishing vacuum expectation value on d-dimensional space-time without destroying the maximal symmetry, then either d or 11-d must equal the number of indices of F, i.e., d=4 or d=7[49*]. Various arguments, though none really convincing, have been advanced to rule out the case of d=7. In fact, a compactification on the four-sphere to seven-dimensional anti-de Sitter space has been produced by Pilch, van Nieuwenhuizen and Townsend.[52] Modulo this d=4, d=11-4 ambiguity, a very important physical parameter, the dimension of space-time, has been predicted. It may be worthwhile to recall the logic of how this comes about. We obtain d=4 as a consequence of F having four indices, which is a consequence of A having three indices, which in turn is a consequence of supersymmetry (84 = 128 - 44). So ultimately it is supersymmetry that determines the dimension of space-time in eleven-dimensional supergravity. There was also a technical assumption concerning F or $^*F$ having a vacuum expectation on space-time. Were one to drop this assumption then further possibilities open up if M_{11-d} is a Kähler manifold.[53] These alternative solutions have not been studied in detail. Even their stability has not yet been clarified.

Below, for obvious reasons, we shall concentrate on the d=4 case. The maximally symmetric form of F is then[49*,54*]

$$F^{MNRS} = \left\{ \begin{array}{ll} \dfrac{\varepsilon^{\mu\nu\rho\sigma}}{\sqrt{|g_4|}} F(x,y) & M,\ldots,S = \mu,\ldots,\sigma \\ F^{mnrs} & , M,\ldots,S = m,\ldots,s \\ 0 & , \text{otherwise} \end{array} \right\} \qquad (4.4)$$

Englert first introduced the F^{mnrs} part which was set equal to zero in reference 49*. In a classical solution the gravitino field is to be set equal to zero. The simplest solutions are obtained by setting h(y)=1 and $F^{mnpq}=0$ in (4.4). Then the field equations and Bianchi identities of eleven-dimensional supergravity require

$$F(x,y) = \mathbf{F} = \text{constant}, \qquad (4.5)$$

and the energy momentum tensor of the A-field is equivalent to two cosmological terms, one on M_4 and one on M_7, with cosmological constants of opposite signs. Provided M_4 contains the time dimension the cosmological constant on M_7 will have the sign appropriate to a compact manifold so that spontaneous compactification really does occur. M_4 is the maximally symmetric non-compact anti-de Sitter (adS) space and M_7 a compact Einstein space. The scale is set by the expectation value **F** of the field strengths. The rest of four-dimensional physics is determined by the shape of the small seven-dimensional manifold M_7. If, for instance, M_7 is the seven-sphere S^7 then the gauge group in four dimensions is SO(8) (the isometry group of S^7) and as pointed out by Duff and Pope[55] one finds eight supersymmetries. The resulting zero mode truncation is[56] the de Wit, Nicolai[57] version of N=8 supergravity. Further compactifications (some with $F^{mnrs}\neq 0$ and/or $h(y)\neq$constant) with assorted gauge symmetries and supersymmetries have been studied by Biran, Englert, de Wit, Nicolai, Casher, Rooman, Awada, Duff, Pope, Nilsson, Castellani, d' Auria, Fré, van Nieuwenhuizen, Gürsey and Tze[58*-64*].

Just as gauge symmetries in four-dimensions are related to the Killing vectors of the small manifold, so the supersymmetries are related to the Killing spinors, as noted by Duff and Pope[55]. These authors also suggested that the transition from the maximally symmetric round seven-sphere to its less symmetric squashed counterpart should correspond to an Englert-Brout-Higgs phenomenon in four dimensions. Spontaneous symmetry breaking and the Englert-Brout-Higgs phenomenon were also explored in the work of d'Auria, Fré and van Nieuwenhuizen[61*]. All the solutions with one or more surviving supersymmetries are stable with respect to classical perturbations[65*]. Some of the solutions without any surviving supersymmetry are stable, and others are unstable. In some cases the spectrum of all higher modes has been worked out.[58*]

Several problems stand in the way of producing a realistic theory. First of all, the four-dimensional adS space has an intolerably large cosmological constant that has to be eliminated somehow. Of course Higgs mechanisms in four dimensions further affect the cosmological constant and it is the end product that has to be very small or zero. Various ideas including that of "foam"[66] have been invoked in this respect, though no clear and workable cosmological constant suppression mechanism is known to us. Another serious problem is the lack of chiral fermions[67*] at least as long as bound states and solitons are ignored.

There is currently great hope that these problems can be solved in the context of higher-dimensional superstrings theory[68,69*,70*,71]. Superstrings also offer the tantalizing possibility of a finite quantum theory of gravity[72*], whereas the eleven-dimensional supergravity theory, while trivially finite[73] at one loop, is questionable in this regard.

1.5 Kaluza-Klein Phenomenology

As noted in the last section, eleven-dimensional space time recommends itself as the habitat of the maximal supergravity theory. Remarkably there is also a phenomenological argument for eleven as the minimal number of space time dimensions, which was pointed out by Witten.[65*] If the familiar $SU(3)_{color} \times (SU(2) \times U(1))_{electroweak}$ gauge symmetry in four dimensions is to originate in isometries of a compact manifold in N "hidden" dimensions, then these extra dimensions must be at least seven in number. This follows from the observation that no manifold of dimension three or smaller can have more than six isometries, and thus the 8-parameter group SU(3) can most economically appear as the isometry group of the four-dimensional manifold CP(2). Similarly an SU(2) x U(1) gauge symmetry in four dimensions is most economically obtained from the isometries of the three-dimensional manifold $S^2 \times S^1$. Thus the gauge group of low energy physics is obtainable from the isometries of the seven-dimensional manifold $CP(2) \times S^2 \times S^1$. This is not an Einstein manifold, however, and as such not relevant as the compactification of eleven-dimensional supergravity. Fortunately coset spaces of the type $SU(3) \times SU(2) \times U(1)/SU(2) \times U(1) \times U(1)$ other than $CP(2) \times S^2 \times S^1$ exist (S^7 is not among them!) and some of these are Einstein manifolds[73]. Alas, because there is no supersymmetric Yang-Mills theory in eleven dimensions, the fermion spectrum is necessarily non-chiral. This will be discussed further in Section 1.8.

1.6 Cosmology

It is natural, in thinking about Kaluza-Klein theories in which the extra dimensions form a compact manifold whose size is not much bigger than the Planck scale, to wonder how the extra dimensions could ever be observed. One area of speculation centers on cosmology: perhaps in the very early universe, when energies approaching the Planck mass were important, the extra dimensions played a direct dynamical role.

The earliest paper to study the relationship of extra dimensions to cosmology was probably that of Forgacs and Horvath[74], who considered a fairly complicated model in six dimensions with gauge fields and matter in the form of dust as well as gravity. For this model, they found numerically a class of solutions compatible with standard Robertson-Walker cosmology.

A much simpler case, namely the vacuum Einstein equations, was considered by Chodos and Detweiler[75*], who pointed out that the five-dimensional Kasner solution[76]

$$ds^2 = -dt^2 + (\frac{t}{t_o})\ (dx^2 + dy^2 + dz^2) + (\frac{t_o}{t})\ (d\omega^2) \quad (6.1)$$

describes a universe in which one spatial dimension shrinks with time while the other three expand. This not only accords with the observed expanding universe, but also can explain why the fifth dimension is so small. (The Einstein equations force at least one of the spatial dimensions to contract; there is no vacuum solution in which all the dimensions expand.)

This idea was generalized to the case of eleven dimensional supergravity by Freund[77*], who found time-dependent versions of the Freund-Rubin compactification[49*] which involve the preferential expansion of three spatial dimensions. The idea has been examined critically by Maeda[78*], who pointed out that particle creation in a time dependent universe might tend to isotropize the anisotropic time development given by the metric of Eq. (6.1).

A second line of development was started by Alvarez and Belén Gavela.[79*] Instead of using cosmology to explain one of the puzzles of Kaluza-Klein theory (why some dimensions are large and others are small) they turned the problem around and asked how the existence of extra dimensions might impact on cosmological issues. In particular, they focussed on the question of entropy production in the early universe, which is intimately connected with the horizon and flatness problems. One answer, of course, is provided by the inflationary universe scenario.[80] Alvarez and Belén Gavela showed that the process of dimensional reduction could be viewed as another mechanism to dump entropy into the observed four dimensional universe. Whether the amount of entropy generated thereby suffices to resolve the flatness problem, however, is not clear. Further discussion of these and related issues have been given by Sahdev[81*] and by Barr and Brown[82].

Other authors have sought to incorporate the mechanism of inflation itself within the context of Kaluza-Klein cosmology. In other words, can one find plausible higher dimensional models which themselves undergo an inflationary phase? Perhaps the first such models were discussed by Dereli and Tucker,[83] who generated the inflation via a suitable higher dimensional energy-momentum tensor, and by Ramond[84]. The idea was examined in greater detail in the context of a higher-derivative model of gravity by Shafi and Wetterich[85*], and by Abbott, Barr and Ellis[86], who suggested that reasonable inflation could be achieved provided that the number of dimensions is large (~ 40).

An important restriction on Kaluza-Klein cosmologies is imposed by the connection between the size of the extra dimensions and the fundamental coupling constants of the theory.[87] The desire to avoid time variation of fundamental constants has led various authors to consider models in which the internal dimensions either remain constant or tend rapidly to an asymptotic limit, even though the observed dimensions continue to evolve. Examples of such models have been given by Ramond[84]; Chapline and Gibbons[88]; Randjbar-Daemi, Salam and Strathdee[89]; Gleiser, Rajpoot and Taylor[90]; Bailin, Love and Vayonakis[91]; and Okada[92].

Another constraint on higher dimensional cosmological models has been addressed by Kolb and Slansky[93*], who analyze the production of the four dimensional particles corresponding to the non-zero modes of the harmonic expansions in mass eigenstates of the higher-dimensional fields. They find that some models can be ruled out, (in particular, the five-dimensional model seems to be inadmissible) and that others are constrained by the requirement that there should be at most a very low density of these heavy particles surviving at the present time.

Recent work by Yoshimura and Koikawa[94] has attempted to study the interplay between quantum effects and cosmology. In particular, by considering the one-loop effective action at finite temperatures in a Robertson- Walker type of background metric, evidence is found for the existence of alternating cycles of expansion and contraction, which they argue may lead to sufficient entropy production to resolve the horizon and flatness problems.

Our understanding of Kaluza-Klein cosmology and its relation to other currently fashionable cosmological ideas is reminiscent of the chaotic and rapidly developing state of the early universe itself. As our knowledge continues to expand,[95] one may hope that some of the key issues that are now being addressed will begin to be decisively resolved.

1.7 Quantum Effects

It was pointed out by Casimir in 1948[96*] that the zero-point fluctuations of the free electromagnetic field give rise to an attractive force between two parallel, infinite, perfectly conducting plates. The presence of the plates imposes boundary conditions on the field, which then modify the zero-point energy in a way that depends on the separation between the plates.

The physics of the Casimir effect generalizes to any situation in which boundary conditions are imposed on a quantum field. In all such cases one expects that observable forces will be generated; they may be either attractive or repulsive. (For example, if the parallel plates are replaced by a thin, perfectly conducting spherical shell, the Casimir force changes sign.[97])

It is clear that a typical Kaluza-Klein theory, with a background manifold of the form $M_4 x B$, will naturally exhibit the Casimir effect because of the restriction imposed by the manifold B. (For example, in the case that B is a circle S^1 of radius ρ, the restriction is just that the momenta are quantized in units of $1/\rho$ (cf.eq. (1.14)). The existence of the Casimir effect in Kaluza-Klein theories was first remarked upon by Unwin[98], although he did not compute it. Duff and Toms[73] discussed the formulation of perturbative quantum gravity in higher dimensions; they noted that if the total number of space-time dimensions is odd then at one loop the theory is finite if a regularization scheme such as dimensional regularization or zeta-function regularization is employed, which does not introduce a dimensionful cutoff.

The Casimir energy in the five-dimensional case was first looked at quantitatively by Appelquist and Chodos[99*] and also by An Ing and Chen Shi[100] in a slightly different context. On purely dimensional grounds, the zero-point energy per unit 3-volume must be of the form

$$V^{(Q)}(\rho) = \frac{C}{\rho^4}$$

where ρ is the radius of the circle and C is a pure number. (Actually, it is proportional to Planck's constant h, which means it can be thought of as induced by one-loop quantum effects.) Because the total space-time dimensionality is odd, C is finite without the need for any subtraction. If one uses a dimensionful cutoff, then one can absorb the cutoff dependence in a redefinition of the cosmological constant. The number C turns out to be $-1/(2\pi)^4 \; 15/4\pi^2$ $\zeta(5)$, where $\zeta(s)$ is the Riemann zeta-function ($\zeta(5) \approx 1.04$). Thus there is an attractive force, just as in the original Casimir effect, which tends to make the extra dimension contract.

This work has been generalized in a variety of ways. Rubin and Barkan Roth[101*] have looked at the same problem in the presence of finite temperature. They find the existence of a critical radius ρ_c, such that for $\rho<\rho_c$ the extra-dimension contracts due to the Casimir effect, whereas for $\rho>\rho_c$, thermal effects make the extra dimension expand. Rubin and Roth[102], and also Tsokos[103], have shown that the addition of massive fermions will produce a repulsive force at short distances that can stabilize the extra dimension; i.e. when considered together with the Casimir energy, the fermion contribution produces a potential with a minimum. Appelquist, Chodos and Myers[104], and also Inami and Yasuda[105], have considered the case of d extra dimensions, compactified into a d-dimensional torus. They find that one of the dimensions will contract, but that the other (d-1) dimensions will expand.

An important step was taken by Candelas and Weinberg[106*], who considered the case in which the internal manifold has non-vanishing curvature. In particular, they considered the set of manifolds $M_4 x S^N$, where M_4 is 4-dimensional Minkowski space, and S^N is the N-sphere. For the reasons alluded to above, N is taken to be odd. They added a bare cosmological term, as well as a set of matter fields (either scalars or spinors). The potential energy is then taken to consist of three terms: (a) the term arising from the curvature of the N-sphere (not present in the case N=1); (b) the cosmological constant term; and (c) the quantum Casimir energy of the matter fields. They neglect the Casimir energy of the gravitational field itself. Their results are, first, that it is often possible to find a stable solution to the equations of motion. Such a solution requires not only that the potential have a minimum as a function of the radius ρ of the sphere, but also that the potential vanish at this minimum. This means there is no net cosmological constant, and therefore the ansatz that the four-dimensional space-time is Minkowski space is a consistent one. Furthermore, stability requires the potential to be repulsive at $\rho=0$, which in turn means that the Casimir force must be opposite in sign to what it was in the case N=1. The fact that this happens is reminiscent of the change of sign in the ordinary electromagnetic Casimir effect in going from parallel plates to a spherical shell.

Perhaps most interesting, the existence of a stable minimum determines the radius of the N-sphere, which in turn, according to standard Kaluza-Klein ideas, fixes the value of the gauge coupling constant[107*] (in the case of an N-sphere, the gauge group is $0(N+1)$). Kaluza-Klein theories are thus possibly unique in offering a dynamical determination of the gauge couplings.

A somewhat disturbing feature of the Candelas-Weinberg work is that the contribution of any one matter field (either scalar or spinor) to the Casimir energy is anomalously small; that is, it contributes to the coefficient of $1/\rho^4$ a number which turns out typically to be of order 10^{-4} or 10^{-5}. As a consequence, if only a small number of matter fields is included, the minimum of the quantum-corrected potential will be much smaller than the Planck length, implying a coupling constant much larger than unity. To remedy this situation, Candelas and Weinberg are forced to contemplate theories with a large (i.e. 10^4 or 10^5) number of matter fields.

A more natural and physical generalization of the five-dimensional study is to consider the contribution of the gravitational field itself to the Casimir energy with a curved internal manifold.[108-110] The computation is much more involved than in the case of matter fields, and there are minor discrepancies among the three groups that have published results. Nevertheless, there is general agreement on the following two points: (i) the gravitational contribution per degree of freedom is much larger than the matter-field contribution; (ii) however, the gravitational field can also contribute an imaginary part to the potential, indicating an instability in the system, and it seems that there is no minimum of the real part of the effective potential in the region where the imaginary part vanishes. (Minima of Re V have been found in cases where $\mathrm{Im}V \neq 0$.) Thus gravity alone at the one-loop level cannot produce a stable spontaneously compactified solution of the form $M^4 \times S^N$ (N odd). It remains to be seen whether a suitable mix of gravitational and matter fields, (such as would be present for example in a supergravity theory; a case discussed, though not computed, by Candelas and Raine[111]) can lead to a different result.

Gibbons and Nicolai[112*], and also Inami and Yamagishi[113*], have examined the question of whether there is a non-vanishing Casimir energy for a supergravity theory expanded about a supersymmetric background, and they conclude, perhaps not surprisingly, that the fermionic and bosonic contributions cancel each other. Thus a classical solution of this type will also be a solution at the one-loop level, since quantum effects generate no change in the potential.

In addition to the work on the Casimir effect, there has been work in a related context by Toms and by Awada and Toms[114], which is higher-dimensional although not, perhaps, strictly in the Kaluza-Klein spirit. They have shown how the one-loop quantum fluctuations of matter fields in higher dimensions can induce the appearance of the gravitational piece of the Lagrangian, similarly to "induced gravity" in four dimensions.[115]

Another exploitation of the higher-dimensional Casimir effect has been made by Hosotani, who has observed that under certain circumstances it can induce the spontaneous breakdown of gauge symmetries.[116]

1.8 The Problem of Chiral Fermions

One of the striking properties of the observed quarks and leptons is that they are chiral; that is, the left-handed and right-handed fermions transform under different representations of the SU(2)xU(1) gauge group. In a Kaluza-Klein theory, these fermions presumably are among the massless modes of some kind of spinor fields in a higher number of dimensions. The question arises as to whether the chiral property descends from the higher-dimensional theory, or whether it is a low-energy manifestation. If the latter, then at some not-too-high energy (i.e. some energy insignificant compared to the compactification scale) mirror partners to the observed fermions will be discovered, such that the combined spectrum of original and mirror fermions is no longer chiral.

There are several arguments against the possibility of mirror fermions, among them: (a) the representations of the observed quarks and leptons repeat themselves in identical families; no mirror partners have yet been discovered. (b) For some miraculous reason, the anomalies that might have infected a chiral theory cancel in the observed spectrum. This removes a potential raison d'être for the mirror fermions, namely that they would necessarily cancel any anomalies present at the chiral level.

The prevalent opinion, then, is that within Kaluza-Klein theories the chiral property derives from a chiral spectrum of fermions in higher dimensions. Witten[65*] was the first to notice that the spectrum of fermions on any eleven-dimensional manifold $M=M_4 x B$, where M_4 is four-dimensional space-time and B is a compact seven-dimensional manifold admitting SU(3)xSU(2)xU(1) symmetry, could not be chiral. Wetterich[117] showed, using the algebraic properties of the gamma-matrices in higher dimensions, that the spectrum of fermions could be chiral only if the dimensionality is 2 mod 8. Finally, using topological arguments, Witten[67*] showed that, in any number of dimensions, the Dirac operator cannot admit a chiral spectrum, and neither can the Rarita-Schwinger operator (with the possible exception of internal manifolds that are not coset spaces). The possible evasions of Witten's results within the context of standard Kaluza-Klein ideas are sufficiently unattractive that some rather unconventional remedies have been proposed.

Perhaps the least radical of these is the suggestion that additional gauge fields be included in the higher-dimensional theory.[118] As shown by Randjbar-Daemi, Salam and Strathdee,[119*] if the compactification involves a topologically non-trivial configuration of these gauge fields, the fermionic spectrum can indeed be chiral. The addition of extra gauge symmetry is precisely the mechanism whereby the superstring theories[120] allow for the appearance of chiral fermions.

Wetterich[121] has pointed out that the proof of Witten's results requires the internal manifold to be compact. He has given examples of non-compact internal manifolds of finite volume, which admit isometries, and also allow chiral fermions. Furthermore, together with Nicolai[122], he has shown that such manifolds can have at least a partially discrete mass spectrum. A further example of the use of non-compact internal manifolds, with relevance to supergravity, has been given by Gell-Mann and Zwiebach.[123*]

Weinberg[124*] has put forward the idea of "quasi-Riemannian geometry" by which he means that in d-dimensions the tangent-space group should be chosen to be something other than O(d-1,1). Thus, for example in the vielbein e^a_μ, the superscript a no longer labels a vector of O(d-1,1) but rather some other representation of some other group. Of course, the tangent-space subgroup corresponding to ordinary space-time must always turn out to be O(3,1). Spinors in a curved manifold actually transform as scalars under general coordinate transformations and spinors under the tangent-space group, so this modification of the tangent-space structure might be expected to wreak havoc with the orderly non-chiral structure of spinors in ordinary Riemannian geometry. Weinberg analyzes the possible tangent space groups, and shows that under the appropriate circumstances the fermionic spectrum can indeed be chiral.

1.9 Monopoles

Given a four-dimensional line element ds_4^2 that satisfies the vacuum Einstein equations, one can obtain a five-dimensional solution by the simple expedient of adding to it $(dx^5)^2$. The line element

$$ds_5^2 = ds_4^2 + (dx^5)^2$$

will automatically be a solution of the five dimensional Einstein equations. This construction, however, is rather trivial and makes no essential use of the extra dimension.

Somewhat less trivial results can be obtained by considering a solution of the four-dimensional **Euclidean** Einstein equation $d\tilde{s}_4^2$, and adding time as the fifth coordinate:

$$d\tilde{s}_5^2 = -dt^2 + d\tilde{s}_4^2.$$

Again this is automatically a solution of the five-dimensional equations, but now the compactification of x^5 will lead to non-trivial restrictions on $d\tilde{s}_4^2$, and upon dimensional reduction, $d\tilde{s}_5^2$ will be interpreted as a static four-dimensional solution involving, in addition to gravity, both electromagnetic and scalar fields. Such solutions have been studied by a number of authors in recent years.[125]

It was noted by Sorkin[126*], and by Gross and Perry[127*], that if $d\tilde{s}_4^2$ is chosen to be the Euclidean Taub-NUT solution then the resulting Kaluza-Klein solution represents a magnetic monopole. The five-dimensional metric is non-singular; the Dirac string singularity emerges as a consequence of dimensional reduction. Gross and Perry also analyze the seemingly strange gravitational behavior of this object arising from the fact that since $g_{00} = -1$, the monopole naively seems to have vanishing gravitational mass. It is possible to generalize the five-dimensional monopole to higher dimensions and higher gauge groups. Perry[128*] has given a topological classification of possible monopoles, and has provided an example of an SO(3) dyon in a six dimensional theory containing gravity and a four-index field strength F_{abcd}.

Other work on spherically-symmetric monopoles has been done by Lee and Lee[129]; quantum effects have been considered by Iwazaki[130] ;and the motion of test particles in the field of a Kaluza-Klein monopole has been investigated by Gegenberg and Kunstatter.[131] A new class of Abelian monopole solutions has been described by Bais and Batenburg.[132]

1.10 **Superstrings**

Higher dimensional unification always includes gravity along with other gauge and matter fields. There being no question as to the need for quantizing these other fields, gravity itself has to undergo quantization as well. Yet no satisfactory quantum theory of gravity exists in four dimensions. No matter how one were to rig up the matter fields, once gravity is included the theory is neither renormalizable nor finite.[133] Despite some early hopes that eleven-dimensional supergravity might be finite to all loops in quantum theory, considerable doubts have arisen as to whether this is in fact so. Green and Schwarz[3*] have made the bold proposal that one radically modify the dynamics of the higher dimensional theory. All theories considered before involved local fields, or in other words point particles, some endowed with internal and Grassmann attributes. Aside from selecting the dimension of space-time, this involves setting to zero the dimensionality δ of the objects hosted by this space time. But what if δ were also open to question? Maybe objects appear point-like but are in reality extended. Extended objects like the nucleon, the hydrogen atom, etc. are of course well known in physics, but invariably they have turned out to be composites of point-like constituents. Yet the new idea that gravity and all other force- and matter-quanta are to be recovered as the excitation modes of a one-dimensional extended object, a string, was first advanced by Scherk and Schwarz[134] in 1975, and then developed by Green and Schwarz.[3*] Strings require a critical dimension[135] for their host space-time and this dimension is then automatically dialed. Nonsupersymmetric strings involve a tachyon and other difficulties. This then leads one unavoidably to consider supersymmetric strings (or superstrings), since supersymmetry in the host space removes any tachyon. The critical dimension is[136] then $d_c=10$. Superstring theories come in four varieties; type I, IIa, IIb[3*] and heterotic.[137,138] The type II theories involve only closed strings and in the pointlike limit as the string tension goes to infinity they reduce to N=2 (non-chiral and chiral) supergravities in ten-dimensions. The heterotic as well as the Type I super-strings involve a gauge group. This gauge group must be SO(32) or E_8xE_8 for the heterotic case, and SO(32) for the type I superstrings with their Chan-Paton rules.[139] For heterotic strings the rank 16 groups E_8xE_8 or SO(32) are arrived at via a toroidal compactification from 26 down to 10 dimensions.[138] For Type I superstrings, the group SO(32) is required for anomaly cancellation. All these (Type I, II and heterotic) theories are anomaly-free, one-loop finite, and for Type IIa and IIb there are now very strong arguments due to Mandelstam[140] that they are finite to all orders. These arguments are expected to generalize to the heterotic case but further techniques are required for Type I superstrings. In any case, meaningful finite quantum theories of gravity could well emerge in this way.

Interesting compactifications to four dimensions with chiral fermions have been described for the E_8xE_8 heterotic superstrings.[68] While it is not our purpose here to describe these theories, we may at least ask: how can it be that they are finite? Their zero mode spectrum is that of N=1 supergravity coupled to an N=1 supersymmetric Yang-Mills system with gauge group E_8xE_8 or SO(32). This theory, though devoid of anomalies, is not finite. How does the superstring cure the divergences? What happens is rather similar to the way that gauge theories cure the non-renormalizable divergences of Fermi's current-current weak-interaction theory. According to Fermi the basic β-decay interaction involves the ultraviolet-dangerous four-fermion vertex (Fig. 1a). In the Glashow-Salam-Weinberg theory this vertex is not fundamental any longer. Rather, the four-Fermi eu → νd process is constructed from the basic (current)-(gauge boson) vertex of the theory, as illustrated in Fig. 1b. This, as is well known, has much better ultra-violet behavior and the ensuing theory is renormalizable.

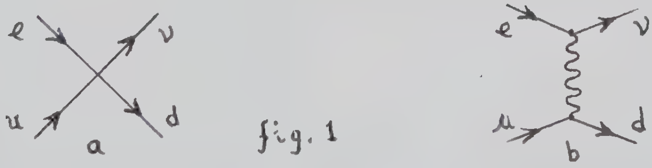


fig. 1

In Einstein's theory there is a three-graviton vertex, a four-graviton vertex, a five-graviton vertex,... and so on (Fig. 2a). The theory is nonpolynomial as the lagrangian contains both the metric matrix and its inverse.

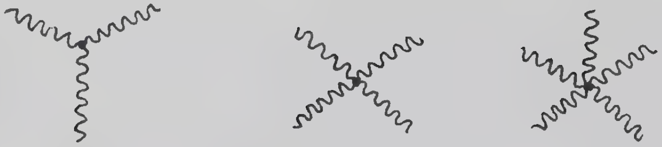


fig. 2a

In the superstring theory, in a particular gauge, there is only a three-superstring vertex. But the superstring has infinitely many excitation modes. If we draw a dotted line for any of the non-zero modes of the superstring, then the three-vertices of Fig. 2b, involving these modes and the graviton, become possible.

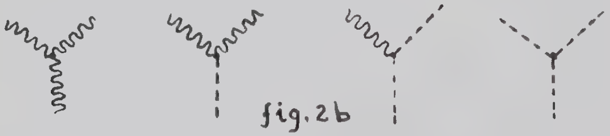


fig. 2b

Now the higher vertices of Fig. 2a can be constructed from the three-vertices of Fig. 2b (summed over all string excitations in virtual lines) much as the four-fermion vertex was constructed from the gauge-theory three-vertices. This is illustrated in Fig. 2c.

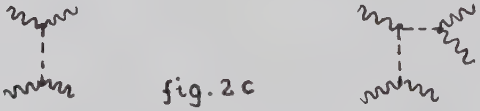


fig. 2c

Again, as in the weak interaction case, a better ultraviolet behavior, and possibly even a **finite** quantum theory ensues.[140] From the electroweak theory, Fermi's nonrenormalizable phenomenological picture reemerges in the singular limit as the W-mass goes to infinity. In the same way here, when the string tension goes to infinity (Scherk limit)[141], the superstring yields a nonrenormalizable limit fraught with infinities: the coupled N=1 supersymmetric Yang-Mills supergravity system.

What all of this teaches us is that aside from the dimension d of space-time, there is a second dimension δ that appears in physics, namely that of the fundamental objects that move in space-time. As long as one sets $\delta=0$, corresponding to point particles, one has a theory of local fields which seemingly cannot accommodate gravity. One cannot help but wonder whether yet further interesting theories[142-144] lie at values of $\delta>1$ (e.g., membranes at $\delta=2$) or whether such theories can somehow be excluded. If such theories exist, then the critical dimensions of their host space-times[136] are of great importance. All these alternatives have to be considered and their consistency checked until hopefully all but one are ruled out. One is dealing here with an ultimate theory of the world and should therefore be entitled to uniqueness.

Historically it is worthwhile noting that the N=1 open and closed superstrings have in the Scherk limit[141] both ten-dimensional Yang-Mills and supergravity sectors. So, what is ultimately reduced to four-dimensions is both a supersymmetric gauge and a supergravity theory, thus supersymmetrically vindicating both the (gauge) Nordström and (gravity) Kaluza-Klein points of view, at least in the Scherk limit. By contrast, the heterotic and Type II superstrings are of the pure Kaluza type.

Finally we wish to point out that when one starts in more than four dimensions (11 or 10) there are a number of mechanisms by which four-dimensional gauge symmetries emerge:

a) from isometries of the small extra-dimensional manifold

b) as hidden symmetries encoded as a σ-model in the scalar fields (e.g., the famous SU(8) of N=8 supergravity), with propagating and composite gauge bosons produced at the quantum level.

c) put in "by hand" in the higher dimension like the SO(32) or $E_8 \times E_8$ of superstring theory needed there to achieve quantum finiteness.

d) as a consequence of approximating extended objects with $\delta>1$ by point particles or strings.[143] e) dynamically, as advocated by H.B. Nielsen et al,[145], by Fradkin and Shenker[146] and by Mandelstam.[147].

Depending on the theory considered, one or more of these gauge-symmetry-producing mechanisms may be at work, and it might be unwise to demand all experimentally observed symmetries to be produced from isometries of the small manifold. It all depends on the dynamics in the higher dimensional space which, as exemplified by superstrings may be of a novel kind.

REFERENCES

1. G. Nordström, Phys. Zeitsch. **13**, 1126 (1912); Ann. d. Physik **40**, 872 (1913); **42**, 533 (1913).

2*. G. Nordström, Phys. Zeitsch. **15**, 504 1914).

3*. M.B. Green and J.H. Schwarz, Phys. Lett. **109B**, 444 (1982); ibid. **149B**, 117 (1984).

4*. Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Klasse 966 (1921).

5. P. Jordan, Ann. d. Physik **1**, 219 (1947).

6. P. Bergmann, Ann. Math. **49**, 255 (1948).

7*. Y. Thiry, Comptes Rendus (Paris) **226**, 216 (1948).

8*. O. Klein, Nature, **118**, 516 (1926).

9. A. Pais, "'Subtle is the Lord...' The Science and the Life of Albert Einstein", Clarendon, Oxford, 1982.

10. S. Chandrasekhar, private communication.

11*. O. Klein, Z.F. Physik **37**, 895 (1926).

12. H. Mandel, Z.F. Physik **39**, 136 (1926).

13. H. Kragh, Alleroed preprint (1983).

14. E. Schroedinger, Ann. d. Phys. **82**, 257 (1927).

15. W. Gordon, Z.F. Physik **40**, 117 (1927).

16. V. Fock, Z.F. Physik **39**, 226 (1926).

17. Th. de Donder and F. van den Dungen, Comptes Rendus (Paris) **183**, 22 (1926).

18. J. Kudar, Ann. d. Phys. **81**, 632 (1926); Phys. Z. **27**, 724 (1926).

19*. A. Einstein and P.G. Bergmann, Ann. Math. **39**, 683 (1938).

20. O. Veblen and B. Hoffmann, Phys. Rev. **36**, 810 (1931).

21. W. Pauli, Ann. d. Physik **18**, 305, 337 (1933).

22. A. Pais, Physica **8**, 1137 (1941).

23. E. Schmutzer, Z.F. Physik **149**, 329 (1957); **154**, 312 (1959).

24. G. Ludwig, "*Fortschritte der Projektiven Relativitätstheorie*" (F. Vieweg und Sohn, Braunschweig 1951).

25. P. Jordan, "Schwerkraft und Weltall" (F. Vieweg und Sohn, Braunschweig, 1955); Z.F. Physik **157**, 112 (1959).

26. M. Fierz, Helv. Phys. Acta **29**, 128 (1956).

27. C. Brans and R. Dicke, Phys. Rev. **124**, 925 (1961).

28. I.B. Rumer "*Cercetari Asupra 5-Opticii*", Editura Tehnica, Bucuresti (1956).

29. J.M. Souriau, Nuovo Cim. **30**, 565 (1963).

30. J. Rayski, Acta Phys. Pol. **27**, 947 (1965) ibid. **28**, 87 (1965).

31. O. Klein in "*New Theories in Physics*" (International Institute of Intellectual Cooperation, Scientific Collection, Paris, 1939) p. 77.

32*. B. DeWitt, in "*Relativity, Groups, Topology*", C. DeWitt and B. DeWitt, eds., (Gordon and Breach, NY, 1964) p. 725.

33. A. Trautman, Rep. Math. Phys. **1**, 29 (1970).

34*. R. Kerner, Ann. Inst. H. Poincaré, **9**, 143 (1968).

35*. Y.M. Cho and P.G.O. Freund, Phys. Rev. **D12**, 1711 (1975).

36. Y.M. Cho, J. Math. Phys. **16,** 2029 (1975).
37*. E. Cremmer and J. Scherk, Nucl. Phys. **B108,** 409 (1976).
38. L. Palla, Proc. 19th International Conference on High-Energy Physics, Tokyo, 1978 (Physical Society of Japan, 1979) p. 629; Z. Horvath, L. Palla, E. Cremmer and J. Scherk, Nucl. Phys. **B127,** 57 (1977).
39*. J.F. Luciani, Nucl. Phys. **B135,** 111 (1978).
40. W. Thirring, Acta. Phys. Austriaca, Suppl. **9,** 256 (1972).
41. G. Domokos and S. Kovesi-Domokos, Nuovo Cimento **44A,** 318 (1978).
42. L.N. Chang, K. MacRae and F. Mansouri, Phys. Rev. **D13** 235 (1976).
43. C.A. Orzalesi, Fortschr. d. Physik **29,** 413 (1981).
44. C. Wetterich, Phys. Lett. **113B,** 377 (1982).
✓ 45*. A. Salam and J. Strathdee, Ann. Phys. (NY) **141,** 316 (1982).
46. L. Dolan and M.J. Duff, Phys. Lett. **52,** 14 (1984).
✓ 47*. E. Cremmer, B. Julia and J. Scherk, Phys. Lett. **76B,** 409 (1978).
48. W. Nahm, Nucl. Phys. **B135,** 149 (1978).
49*. P.G.O. Freund and M.A. Rubin, Phys. Lett. **97B,** 233 (1980).
50. B. de Wit and H. Nicolai, Phys. Lett. **148B,** 60 (1984).
51. P. van Nieuwenhuizen in "*Relativity, Groups and Topology II*", B. de Wit and R. Stora, eds. (North Holland, Amsterdam, 1984).
52. K. Pilch, P. Townsend and P. van Nieuwenhuizen, Nucl. Phys. **B242,** 377 (1984).
53. B. Dolan, Phys. Lett. **140B,** 304 (1984).
54*. F. Englert, Phys. Lett. **119B,** 339 (1982).
55. M.J. Duff, Nucl. Phys. **B219,** 389 (1983).

56. B. de Wit, H. Nicolai and N.P. Warner, CERN Preprint CERN-TH 4052/84 (1984).

57. B. de Wit and H. Nicolai, Phys. Lett. **108B**, 285 (1981).

58*. B. Biran, F. Englert, B. de Wit and H. Nicolai, Phys. Lett. **124B**, 45 (1983).

59*. A. Casher, F. Englert, H. Nicolai and M. Rooman, Nucl. Phys. **B243**, 173 (1984).

60*. M.A. Awada, M.J. Duff and C.N. Pope, Phys. Rev. Lett. **50**, 294 (1983).

61*. M.J. Duff, B.E.W. Nilsson and C.N. Pope, Phys. Rev. Lett. **50**, 2043 (1983); **51**, 846(E) (1983).

62*. R. d'Auria, P. Fré, and P. van Nieuwenhuizen, Phys. Lett. **122B**, 225 (1983).

63*. F. Gürsey and H.C. Tze, Phys. Lett. **127B**, 191 (1983).

64. L. Castellani, R. d'Auria and P. Fré, Nucl. Phys. **B239**, 610 (1984).

65*. E. Witten, Nucl. Phys. **B186**, 412 (1981).

66. S.W. Hawking, Nucl. Phys. **B144**, 349 (1978).

67*. E. Witten, in "Proceedings of the Shelter Island Conference" (MIT Press, Cambridge, Mass. (1985)).

68. P. Candelas, G. Horowitz, A. Strominger and E. Witten, Santa Barbara Preprint NSF-ITP-84-170 (1984).

69*. G. Chapline and N. Manton, Phys. Lett. **120B**, 105 (1983).

70*. D.Z. Freedman, G.W. Gibbons and P.C. West, Phys. Lett. **124B**, 491 (1983).

71. S. Randjbar-Daemi, A. Salam and J. Strathdee, Phys. Lett. **124B**, 349 (1983).

72*. M.B. Green and J.H. Schwarz, Phys. Lett. **149B**, 117 (1984).

73. M.J. Duff and D.J. Toms in "Unification of the Fundamental Interactions II", S. Ferrara and J. Ellis, eds. (Plenum, New York (1983)) p. 29.

74. P. Forgacs and Z. Horvath, Gen. Rel. and Grav. **11** , 205 (1979).

75*. A. Chodos and S. Detweiler, Phys. Rev. **D21**, 2167 (1980).

76. E. Kasner, Am. J. Math. **48**, 217 (1921).

77*. P.G.O. Freund, Nucl. Phys. **B209**, 146 (1982).

78*. K. Maeda, Phys. Lett. **138B**, 269 (1984). (See also T. Koikawa and K. Maeda, Phys. Lett. **149B** 82 (1984).)

79*. E. Alvarez and M. Belén Gavela, Phys. Rev. Lett. **51**, 931 (1983).

80. A.H. Guth, Phys. Rev. **D23**, 348 (1982); A.D. Linde, Phys. Lett. **108B**, 389 (1984); A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

81*. D. Sahdev, Phys. Lett. **137B**, 155 (1984); Phys. Rev. **D30**, 2495 (1985).82. S. Barr and L. Brown, Phys. Rev. **D29**, 2779 (1984).

82. S. Barr and L. Brown, Phys. Rev. **D29**, 2779 (1984).

83. T. Dereli and R.W. Tucker, Phys. Lett. **125B**, 133 (1983).

84. P. Ramond, in Lattice Gauge Theories - Supersymmetry and Grand Unification (The Johns Hopkins Univ., 1982) p. 223.

85*. Q. Shafi and C. Wetterich, Phys. Lett. **129B**, 387 (1983).

86. R. Abbott, S.M. Barr and S.D. Ellis, Phys. Rev. **D30**, 720 (1984).

87. W. Marciano, Phys. Rev. Lett. **52**, 489 (1984).

88. G.F. Chapline and G.W. Gibbons, Phys. Lett. **135B**, 43 (1984).

89*. S. Randjbar-Daemi, A. Salam and J. Strathdee, Phys. Lett. **135B**, 388 (1984).

90. M. Gleiser, S. Rajpoot and J.G. Taylor, Phys. Rev. **D31**, 1904 (1985).

91. D. Bailin, A. Love and C. Vayonakis, Phys. Lett. **142B**, 344 (1984).

92. Y. Okada, Phys. Lett. **150B**, 103 (1985).

93*. E.W. Kolb and R. Slansky, Phys. Lett. **135B**, 378 (1984).

94. M. Yoshimura, Phys. Rev. **D30**, 344 (1984); M. Yoshimura and T. Koikawa, Phys. Lett. **150B**, 107 (1985), and Phys. Lett. **155B**, 241 (1985).

95. Some other interesting papers include: R. Bergamini and C. Orzalesi, Phys. Lett. **135B**, 38 (1984); K. Maeda and H. Nishino, preprint IC/84/169, ICTP, Trieste, (1984); M.A. Rubin, preprint UTTG-15-84, Univ. of Texas, (1984), to appear in Inner Space/Outer Space, ed. by E.W. Kolb, et al., (Chicago Univ. Press, 1985).

96*. H.B.G. Casimir, Proc. Kon. Ned. Akad. v. Wet. **51**, 793 (1948).

97. T.H. Boyer, Phys. Rev. **174**, 1764 (1968).

98. S. Unwin, Phys. Lett. **103B**, 18 (1981).

99*. T. Appelquist and A. Chodos, Phys. Rev. Lett. **50**, 141 (1983); Phys. Rev. **D28**, 772 (1983).

100. An Ing and Chen Shi, Proceedings of the Third M. Grossmann Meeting in General Relativity Part B (Shanghai, 1982) p. 1091.

101*. M.A. Rubin and B. Barkan-Roth, Nucl. Phys. **B226**, 444 (1983).

102. M.A. Rubin and B. Roth, Phys. Lett. **127B**, 55 (1983).

103. K. Tsokos, Phys. Lett. **126B**, 451 (1983).

104. T. Appelquist, A. Chodos and E. Myers, Phys. Lett. **127B**, 51 (1983).

105. T. Inami and O. Yasuda, Phys. Lett. **133B**, 180 (1983).

106*. P. Candelas and S. Weinberg, Nucl. Phys. **B237,** 397 (1984); see also Ya. I. Kogan and N.A. Voronov, JETP Letters, **38,** 311 (1983).

107*. S. Weinberg, Phys. Lett. **125B,** 265 (1983).

108. M.H. Sarmadi, preprint IC/84/3, ICTP, (Jan. 1984).

109. M.A. Rubin and C.R. Ordoñez, J. Math. Phys. **25,** 2888 (1984); J. Math. Phys. **26,** 65 (1985); C. Ordoñez and M.A. Rubin, preprint UTTG-18-84, Univ. of Texas, (December, 1984).

110. A. Chodos and E. Myers, Ann. of Phys. **156,** 412 (1984); Proceedings of the Johns Hopkins Workshop 8, The Johns Hopkins Univ., (1984); Brookhaven preprint BNL-35799, (Jan. 1985), to appear in Phys. Rev. D.

111. P. Candelas and D.J. Raine, Nucl. Phys. **B243,** 157 (1984).

112*. G. Gibbons and H. Nicolai, Phys. Lett. **143B,** 108 (1984).

113*. T. Inami and K. Yamagishi, Phys. Lett. **143B,** 115 (1984).

114. D.J. Toms, Phys. Lett. **129B,** 31 (1983); M. Awada and D.J. Toms, Nucl. Phys. **B245,** 161 (1984); Phys. Lett. **135B,** 283 (1984).

115. See, e.g., A.D. Sakharov, Sov. Phys. Doklady, **12,** 54 (1968); S.L. Adler, Rev. Mod. Phys. **54,** 54 (1982), Erratum ibid. **55,** 837 (1983).

116. Y. Hosotani, Phys. Lett. **129B,** 193 (1983); Phys. Rev. **D29,** 731 (1984).

117. C. Wetterich, Nucl. Phys. **B222,** 20 (1983).

118. Z. Horvath, L. Palla, E. Cremmer and J. Scherk, Nucl. Phys. **B127,** 57 (1977).

119*. S. Randjbar-Daemi, A Salam and J. Strathdee, Nucl. Phys. **B214,** 491 (1983); see also N.S. Manton, Nucl. Phys. **B193,** 502 (1981).

120. M.B. Green and J.H. Schwarz, Phys. Lett. **B151**, 21 (1985); Phys. Lett. **B149**, 117 (1984); D.J. Gross, J.A. Harvey, E. Martinec and R. Rohm, Phys. Rev. Lett. **54**, 502 (1985).

121. C. Wetterich, Nucl. Phys. **B242**, 473 (1984); Nucl. Phys. **B244**, 359 (1984); Nucl. Phys. **B253**, 366 (1985); preprint BUTP-1985/4, Univ. of Bern, (1985).

122. H. Nicolai and C. Wetterich, Phys. Lett. **150B**, 347 (1985).

123*. M. Gell-Mann and B. Zwiebach, Phys. Lett. **147B**, 111 (1984).

124*. S. Weinberg, Phys. Lett. **138B**, 47 (1984); Fifth Workshop on Grand Unification, ed. by K. Kang, H. Fried and P. Frampton, (World Scientific, 1984) p. 508.

125. A. Chodos and S. Detweiler, Gen. Rel. and Grav. **14**, 879 (1982); V. Belinsky and R. Fuffini, Phys. Lett. **89B**, 195 (1980); J. Burzlaff and D. Maison, Journ. Math. Phys. **20**, 1352 (1979).

126*. R. Sorkin, Phys. Rev. Lett. **51**, 87 (1983).

127*. D.J. Gross and M.J. Perry, Nucl. Phys. **B226**. 29 (1983).

128*. M.J. Perry, Phys. Lett. **137B**, 171 (1984).

129. H.-M. Lee and S.-C. Lee, Phys. Lett. **149B**, 95 (1984).

130. A. Iwazaki, Progr. Theor. Phys. **72**, 834 (1984).

131. J. Gegenberg and G. Kunstatter, Phys. Lett. **106A**, 410 (1984).

132. F.A. Bais and P. Batenburg, Nucl. Phys. **B253**, 162 (1985).

133. S. Deser and P. van Nieuwenhuizen, Phys. Rev. Lett. **32**, 245 (1974); Phys. Rev. **D10**, 401 (1974); ibid, 411 (1974).

134. J. Scherk and J.H. Schwarz, Nucl. Phys. **B81**, 118 (1974).

135. C. Lovelace, Phys. Lett. **34B,** 500 (1971).
136. P. Goddard and R.E. Waltz, Nucl. Phys. **B34,** 99 (1971).
137. D. Gross, J. Harvey, E. Martinec and R. Rohm, Phys. Lett. **54,** 502 (1985).
138. P.G.O. Freund, Phys. Lett. **151B,** 387 (1985).
139. J.E. Paton and H.M. Chan, Nucl. Phys. **B10,** 519 (1969).
140. S. Mandelstam (private communication).
141. J. Scherk, Nucl. Phys. **B31,** 222 (1971).
142. A. Sugamoto, Nucl.. Phys. **B215,** 381 (1983); P.A. Collins and R.W. Tucker, Nucl. Phys. **B112,** 159 (1976); P.S. Howe and R.W. Tucker, J. Math. Phys. **19,** 869, 981 (1978).
143. P.G.O. Freund and F. Mansouri, Z.F. Physik **C14,** 279 (1982).
144. S. Randjbar-Daemi, A. Salam, E. Sezgin and J. Strathdee, Phys. Lett. **151B,** 351 (1985).
145. D. Foerster, H.B. Nielsen and M. Ninomiya, Phys. Lett. **94B,** 135 (1980); S. Chadha and H.B. Nielsen, Nucl. Phys. **B217,** 125 (1983).
146. E. Fradkin and S. Shenker, Phys. Rev. **D19,** 3682 (1979).
147. S. Mandelstam (unpublished).

CHAPTER
2
History

On the Possibility of a Unification of the Electromagnetic and Gravitation Fields

by Gunnar Nordström

One of the great merits of relativity theory is of course its ability to characterize the electromagnetic state of the aether through one vector, Minkowski's six-vector f, whereas in the old interpretation two field-vectors were needed. This ability, to characterize the aether state through one vector, is lost, as soon as one assumes besides the electromagnetic field, also a gravitational field in the aether. In the gravity theories developed by Herr Mie[1] and by me,[2] the gravitational field in the aether is given by a four vector; if such a theory is realistic, then the state of the aether is to be described by a six-vector and a four vector.

We shall denote by

$$f_{xy}, f_{yz}, f_{zx}, f_{xu}, f_{yu}, f_{zu}$$

the components of the six-vector, where $u = cit$ and c is the velocity of light. For the components of the magnetic field strength H and of the electric field-strength E one has[3]

$$\begin{aligned} H_x &= f_{yz} = -f_{zy}, \text{ etc...} \\ -iE_x &= f_{xu} = =f_{ux}, \text{ etc...} \end{aligned} \tag{1}$$

Purely formally, we introduce the following notation for the components of the four-vector describing the gravitational field:

$$f_{wx}, f_{wy}, f_{wz}\ f_{wu},$$

---

[1] G. Mie, Ann. Physik **40**, 25, 1913.

[2] G. Nordström, Physik. Zeitschr. **13** 1126, 1912; Ann. d. Phys. **40**, 872, 1913; **42**, 533, 1913.

[3] H. Minkowski, Gött. Nachr. 1903, p. 58.

(where a $f_{xw} = -f_{wx}$, etc...). We then write the following system of equations:

$$\begin{aligned}
&\frac{\partial f_{xy}}{\partial y} + \frac{\partial f_{xz}}{\partial z} + \frac{\partial f_{xu}}{\partial u} + \frac{\partial f_{xw}}{\partial w} = \frac{1}{c}\, k_x,\\
\frac{\partial f_{yx}}{\partial x} &\qquad\qquad\;\; \frac{\partial f_{yz}}{\partial z} + \frac{\partial f_{yu}}{\partial u} + \frac{\partial f_{yw}}{\partial w} = \frac{1}{c}\, k_y,\\
\frac{\partial f_{zx}}{\partial x} &+ \frac{\partial f_{zy}}{\partial y} \qquad\qquad + \frac{\partial f_{zu}}{\partial u} + \frac{\partial f_{zw}}{\partial w} = \frac{1}{c}\, k_z,\\
\frac{\partial f_{ux}}{\partial x} &+ \frac{\partial f_{uy}}{\partial y} + \frac{\partial f_{uz}}{\partial z} \qquad\qquad + \frac{\partial f_{uw}}{\partial w} = \frac{1}{c}\, k_u,\\
\frac{\partial f_{wx}}{\partial x} &+ \frac{\partial f_{wy}}{\partial y} + \frac{\partial f_{wz}}{\partial z} + \frac{\partial f_{wu}}{\partial u} \qquad\qquad = \frac{1}{c}\, k_w,
\end{aligned} \tag{I}$$

$$\begin{aligned}
&\frac{\partial f_{yz}}{\partial x} + \frac{\partial f_{zx}}{\partial y} + \frac{\partial f_{xy}}{\partial z} = 0,\\
&\frac{\partial f_{zu}}{\partial y} + \frac{\partial f_{uy}}{\partial z} + \frac{\partial f_{yz}}{\partial u} = 0,\\
&\frac{\partial f_{xu}}{\partial z} + \frac{\partial f_{uz}}{\partial x} + \frac{\partial f_{zx}}{\partial u} = 0,\\
&\frac{\partial f_{yu}}{\partial x} + \frac{\partial f_{ux}}{\partial y} + \frac{\partial f_{xy}}{\partial u} = 0,\\
&\frac{\partial f_{zw}}{\partial y} + \frac{\partial f_{wy}}{\partial z} + \frac{\partial f_{yz}}{\partial w} = 0,\\
&\frac{\partial f_{xw}}{\partial z} + \frac{\partial f_{wz}}{\partial x} + \frac{\partial f_{zx}}{\partial w} = 0,\\
&\frac{\partial f_{yw}}{\partial x} + \frac{\partial f_{wx}}{\partial y} + \frac{\partial f_{xy}}{\partial w} = 0,\\
&\frac{\partial f_{uw}}{\partial x} + \frac{\partial f_{wx}}{\partial u} + \frac{\partial f_{xu}}{\partial w} = 0,\\
&\frac{\partial f_{uw}}{\partial y} + \frac{\partial f_{wy}}{\partial u} + \frac{\partial f_{yu}}{\partial w} = 0,\\
&\frac{\partial f_{uw}}{\partial z} + \frac{\partial f_{wz}}{\partial u} + \frac{\partial f_{zu}}{\partial w} = 0.
\end{aligned} \tag{II}$$

Both systems of equations are completely symmetric with respect to x, y, u, v, w. As they stand, they have no physical meaning. Should one however set all partial derivatives with respect to w equal to zero, one would then find that these equations become precisely the field equations of the electromagnetic and gravitational fields, with k_x, k_y, k_z, k_u the components of the four current, and $-\frac{1}{c}k_w$ the density at rest of the gravitating mass.* The first four equations in (I) and (II) are Maxwell's equations in the form given by Minkowski. The last equation (I) is the fundamental equation of gravity, and the six remaining equations (II) express the fact that the gravitational vector has vanishing curl.

This interpretation of equations (I) and (II) shows that we are entitled to regard the four-dimensional space-time as a surface in a five-dimensional world. In that five-dimensional world, the k_m are the components of a five-vector, and the f_{mn} the components of a ten-vector. The latter completely characterizes the aether-state. The five-dimensional world has a preferred axis, the w axis. The four-dimensional world is orthogonal to this axis, and in all its points, the derivatives wtih respect to w of all components of f vanish.

The components of f can be expresed in terms of a five-potential with components ϕ_x, ϕ_y, ϕ_z, ϕ_u, ϕ_w

$$f_{mn} = \frac{\partial \phi_n}{\partial m} - \frac{\partial \phi_m}{\partial n} \tag{2}$$

Differentiating the equations (I), one finds for the "five-current" the relation

$$\frac{\partial k_x}{\partial x} + \frac{\partial k_y}{\partial y} + \frac{\partial k_z}{\partial z} + \frac{\partial k_u}{\partial u} + \frac{\partial k_w}{\partial w} = 0, \tag{3}$$

and therefore one can impose the following six partial differential equations for the five-potential

$$\frac{\partial \phi_x}{\partial x} + \frac{\partial \phi_y}{\partial y} + \frac{\partial \phi_z}{\partial z} + \frac{\partial \phi_u}{\partial u} + \frac{\partial \phi_w}{\partial w} = 0 \tag{4}$$

---

* By this I mean the quantity which I called $g \cdot \nu$ in the quoted papers: $-\frac{1}{c}k_w = g \cdot \nu$. More generally one could introduce the notations af_{wx}, af_{wy}, af_{wz} af_{wu} for the components of the gravitational vector, with a an arbitrary real or imaginary constant. Then one would have $-\frac{a}{c}f_w = g\nu$. When deriving the theorem of energy-momentum conservation, it will be shown that a must be $+1$ or -1 in order for the last equation (I) to really behave as the others.

$$\frac{\partial^2\phi_x}{\partial x^2} + \frac{\partial^2\phi_x}{\partial y^2} + \frac{\partial^2\phi_x}{\partial z^2} + \frac{\partial^2\phi_x}{\partial u^2} + \frac{\partial^2\phi_x}{\partial w^2} = - \frac{1}{c}\, k_x,$$

$$\frac{\partial^2\phi_w}{\partial x^2} + \frac{\partial^2\phi_w}{\partial y^2} + \frac{\partial^2\phi_w}{\partial z^2} + \frac{\partial^2\phi_w}{\partial u^2} + \frac{\partial^3\phi_w}{\partial w^2} == - \frac{1}{c}\, k_w. \tag{5}$$

We only had space here, to write out in detail the first and last of the five equations (5). When these six conditions are met, the expressions (2) identically obey the field-equations (I), (II).

Our formulae (I), (II) allow us to write down energy-momentum conservation in a unified way for the combined electromagnetic and gravitational fields. To obtain momentum conservation in the x-direction, multiply those four equations (I) which refer to the other axes by $f_{xy}, f_{xz}, f_{xu}, f_{xw}$, and those six equations (II) which contain x by f_{mn} where m, n are the properly ordered remaining two indices appearing in the corresponding equation. Add the so obtained ten equations, and some simple manipulations then yield the desired result.

We want to explicitly establish the theorem for the u-direction - i.e., energy conservation - and therefore multiply the first three equations (I) by $f_{ux}, f_{uy}f_{uz}$. and the last one by f_{uw}. Of the equations (II), the second, third, fourth, eighth, ninth and tenth are to be multiplied by $f_{yz}, f_{zx}, f_{xy}, f_{wx}, f_{wy}, f_{wz}$ respectively. After adding, and rearranging some terms we obtain

$$f_{ux}\left(\frac{\partial f_{xy}}{\partial y} + \frac{\partial f_{xz}}{\partial z} + \frac{\partial f_{xw}}{\partial w}\right) +$$

$$+ f_{uy}\left(\frac{\partial f_{yx}}{\partial x} + \frac{\partial f_{yz}}{\partial z} + \frac{\partial f_{yw}}{\partial w}\right) +$$

$$+ f_{uz}\left(\frac{\partial f_{zx}}{\partial x} + \frac{\partial f_{zy}}{\partial y} + \frac{\partial f_{zw}}{\partial w}\right) +$$

$$+ f_{uw}\left(\frac{\partial f_{wx}}{\partial x} + \frac{\partial f_{wy}}{\partial y} + \frac{\partial f_{wz}}{\partial z}\right) +$$

$$+ f_{yz}\left(\frac{\partial f_{zu}}{\partial y} + \frac{\partial f_{uy}}{\partial z}\right) + f_{zx}\left(\frac{\partial f_{xu}}{\partial z} + \frac{\partial f_{uz}}{\partial z}\right) +$$

$$
\begin{aligned}
&+ f_{xy}\left(\frac{\partial f_{yu}}{\partial x} + \frac{\partial f_{ux}}{\partial y}\right) + f_{wx}\left(\frac{\partial f_{uw}}{\partial x} + \frac{\partial f_{xu}}{\partial w}\right) + \\
&+ f_{wy}\left(\frac{\partial f_{uw}}{\partial y} + \frac{\partial f_{yu}}{\partial w}\right) + f_{wz}\left(\frac{\partial f_{uw}}{\partial z} + \frac{\partial f_{zu}}{\partial w}\right) + \\
&+ f_{ux}\left(\frac{\partial f_{xu}}{\partial u} + f_{uy}\frac{\partial f_{yu}}{\partial u} + f_{uz}\frac{\partial f_{zu}}{\partial u} + f_{uw}\frac{\partial f_{wu}}{\partial u}\right) - \\
&+ f_{yz}\frac{\partial f_{yz}}{\partial u} + f_{zx}\frac{\partial f_{zx}}{\partial u} + f_{xy}\frac{\partial f_{xy}}{\partial u} + \\
&+ f_{wx}\frac{\partial f_{wx}}{\partial u} + f_{wy}\frac{\partial f_{wy}}{\partial u} + f_{wz}\frac{\partial f_{wz}}{\partial u} = \\
&= \frac{1}{c}(f_{ux}k_x + f_{uy}k_y + f_{uz}k_z + f_{uw}k_w).
\end{aligned}
$$

A simple manipulation yields the desired equation

$$
\begin{aligned}
&\frac{\partial}{\partial x}(f_{uy}f_{yx} + f_{uz}f_{zx} + f_{uw}f_{wx}) + \\
&+ \frac{\partial}{\partial y}(f_{ux}f_{xy} + +) + \frac{\partial}{\partial z}(f_{ux}f_{xz} + +) + \\
&+ \frac{\partial}{\partial w}(f_{ux}f_{xw} + +) + \frac{1}{2}\frac{\partial}{\partial u}(- f_{ux}^2 - \\
&- f_{uy}^2 - f_{uz}^2 - f_{uw}^2 + f_{yz}^2 + f_{zx}^2 + \\
&+ f_{xy}^2 + f_{wx}^2 + f_{wy}^2 + f_{wz}^2) = \\
&= \frac{1}{c}(f_{ux}k_x + f_{uy}k_y + f_{uz}k_z + f_{uw}k_w).
\end{aligned} \tag{6}
$$

The quantities in parantheses on the left-hand side, are the components of a five-dimensional tensor and consequently the equation has the form

$$
\frac{\partial P_{ux}}{\partial x} + \frac{\partial P_{uy}}{\partial y} + \frac{\partial P_{uz}}{\partial z} + \frac{\partial P_{uu}}{\partial u} + \frac{\partial P_{uw}}{\partial w} = K_u. \tag{6a}
$$

Multiplying by ic and setting $\frac{\partial P_{uw}}{\partial w}$ to zero, one finds the usual form of the energy conservation law. The x component of the energy-current is given by

$$\begin{aligned} S_x &= ic(f_{uy}f_{yx} + f_{uz}f_{zx} + f_{uw}f_{wx}) \\ &= c(E_yH_z - E_zH_y) - \frac{\partial \phi_w}{\partial t}\frac{\partial \phi_w}{\partial x} \; . \end{aligned} \tag{7}$$

The 1st form is obtained using equations (1) and (2), and the appearing w derivatives have been set to zero. In vector notation the expression for S is

$$S = c[EH] - \frac{\partial \phi_w}{\partial t} \nabla \phi_w \; . \tag{7a}$$

from (6) one also obtains an expression for the energy density, which again invector notation is

$$\psi = \frac{1}{2}\left[E^2 + H^2 + (\nabla \phi_w)^2 + \frac{1}{c^2}\left(\frac{\partial \phi_w}{\partial t}\right)^2\right] .$$

These expressions for the energy-current and energy-density are formed additively from the corresponding electromagnetic and gravitational expressions. This is the desired result of our considerations.

One readily notices the validity of the statement in the footnote * above, so that of the components of the ten-vector f and of the five-vector k only those that have a u-index are imaginary.

Permuting the indices in equation (6), one finds four further equations. The three referring to space-like directions express the known laws of momentum conservation. The equation for the w-direction, expressed in terms of the electromagnetic field strengths and of the gravitational potential ϕ_w, is in vector notation

$$\begin{aligned}
&- \operatorname{div}\left[[H, \nabla \phi_w] + \frac{1}{c} E \frac{\partial \phi_w}{\partial t}\right] + \frac{1}{c} E \frac{\partial}{\partial t}(E \nabla \phi_w) + \\
&+ \frac{1}{2} \frac{\partial}{\partial w}\left[H^2 - E^2 - (\nabla \phi_w)^2 + \frac{1}{c^2}\left(\frac{\partial \phi_w}{\partial t}\right)^2\right] = \\
&= - \frac{1}{c}\left[k \nabla \phi_w + k_u \frac{\partial \phi_w}{\partial u}\right].
\end{aligned}$$

For the time being it is not clear what physical meaning is to be attributed to this equation.

The above approach offers formal advantages, in that the electromagnetic and gravitational fields can be regarded as a single field. No new physical content is thereby given to the equations. I still would not exclude a possible deeper origin for the formal symmetry we found. Here however, I shall not go into the possibilities which one can think of in this respect.

Abstract. It is shown that a unified treatment of the electromagnetic and gravitational fields is possible if one views the four-dimensional space time as a surface in a five-dimensional world.

Helsingfors, March 30, 1914

(received April 3, 1914).

Translated from the German Original by P.G.O.F.

Über die Möglichkeit, das elektromagnetische Feld und das Gravitationsfeld zu vereinigen.

Von Gunnar Nordström.

Es ist ja eins der großen Verdienste der Relativitätstheorie, daß sie den elektromagnetischen Zustand des Äthers durch einen Vektor, den Minkowskischen Sechservektor $\mathfrak{f}$ zu charakterisieren vermag, während nach der alten Auffassung hierfür zwei Feldvektoren erforderlich waren. Diese Möglichkeit, den Ätherzustand durch einen Vektor zu charakterisieren, wird aber hinfällig, sobald man außer dem elektromagnetischen Felde noch ein Gravitationsfeld im Äther annimmt. In den Gravitationstheorien, die Herr Mie[1]) und ich[2]) entwickelt haben, wird das Gravitationsfeld im Äther durch einen Vierervektor angegeben; wenn eine derartige Theorie der Wahrheit entspricht, wird also der Ätherzustand durch einen Sechservektor und einen Vierervektor charakterisiert.

Wir wollen die Komponenten des elektromagnetischen Sechservektors mit

$$\mathfrak{f}_{xy},\ \mathfrak{f}_{yz},\ \mathfrak{f}_{zx},\ \mathfrak{f}_{xu},\ \mathfrak{f}_{yu},\ \mathfrak{f}_{zu}$$

bezeichnen, indem $u = cit$ gesetzt ist, wo c die Lichtgeschwindigkeit. Für die Komponenten der magnetischen Feldstärke $\mathfrak{H}$ und der elektrischen Feldstärke $\mathfrak{E}$ hat man dann[3])

$$\left.\begin{aligned} \mathfrak{H}_x &= \mathfrak{f}_{yz} = -\mathfrak{f}_{zy} \quad \text{usw.,} \\ -i\mathfrak{E}_x &= \mathfrak{f}_{xu} = -\mathfrak{f}_{ux} \quad \text{usw.} \end{aligned}\right\} \qquad (1)$$

Rein formal führen wir nun weiter für die Komponenten des Vierervektors des Gravitationsfeldes folgende Bezeichnungen ein:

$$\mathfrak{f}_{wx},\ \mathfrak{f}_{wy},\ \mathfrak{f}_{wz},\ \mathfrak{f}_{wu},$$

(wobei $\mathfrak{f}_{xw} = -\mathfrak{f}_{wx}$ usw.) und schreiben folgende Gleichungssysteme auf:

$$\left.\begin{aligned}
\frac{\partial \mathfrak{f}_{xy}}{\partial y} + \frac{\partial \mathfrak{f}_{xz}}{\partial z} + \frac{\partial \mathfrak{f}_{xu}}{\partial u} + \frac{\partial \mathfrak{f}_{xw}}{\partial w} &= \frac{1}{c}\,\mathfrak{s}_x, \\
\frac{\partial \mathfrak{f}_{yx}}{\partial x} + \frac{\partial \mathfrak{f}_{yz}}{\partial z} + \frac{\partial \mathfrak{f}_{yu}}{\partial u} + \frac{\partial \mathfrak{f}_{yw}}{\partial w} &= \frac{1}{c}\,\mathfrak{s}_y, \\
\frac{\partial \mathfrak{f}_{zx}}{\partial x} + \frac{\partial \mathfrak{f}_{zy}}{\partial y} + \frac{\partial \mathfrak{f}_{zu}}{\partial u} + \frac{\partial \mathfrak{f}_{zw}}{\partial w} &= \frac{1}{c}\,\mathfrak{s}_z, \\
\frac{\partial \mathfrak{f}_{ux}}{\partial x} + \frac{\partial \mathfrak{f}_{uy}}{\partial y} + \frac{\partial \mathfrak{f}_{uz}}{\partial z} + \frac{\partial \mathfrak{f}_{uw}}{\partial w} &= \frac{1}{c}\,\mathfrak{s}_u, \\
\frac{\partial \mathfrak{f}_{wx}}{\partial x} + \frac{\partial \mathfrak{f}_{wy}}{\partial y} + \frac{\partial \mathfrak{f}_{wz}}{\partial z} + \frac{\partial \mathfrak{f}_{wu}}{\partial u} &= \frac{1}{c}\,\mathfrak{s}_w,
\end{aligned}\right\} \qquad \text{(I)}$$

$$\left.\begin{aligned}
\frac{\partial \mathfrak{f}_{yz}}{\partial x} + \frac{\partial \mathfrak{f}_{zx}}{\partial y} + \frac{\partial \mathfrak{f}_{xy}}{\partial z} &= 0, \\
\frac{\partial \mathfrak{f}_{zu}}{\partial y} + \frac{\partial \mathfrak{f}_{uy}}{\partial z} + \frac{\partial \mathfrak{f}_{yz}}{\partial u} &= 0, \\
\frac{\partial \mathfrak{f}_{xu}}{\partial z} + \frac{\partial \mathfrak{f}_{uz}}{\partial x} + \frac{\partial \mathfrak{f}_{zx}}{\partial u} &= 0, \\
\frac{\partial \mathfrak{f}_{yu}}{\partial x} + \frac{\partial \mathfrak{f}_{ux}}{\partial y} + \frac{\partial \mathfrak{f}_{xy}}{\partial u} &= 0, \\
\frac{\partial \mathfrak{f}_{zw}}{\partial y} + \frac{\partial \mathfrak{f}_{wy}}{\partial z} + \frac{\partial \mathfrak{f}_{yz}}{\partial w} &= 0, \\
\frac{\partial \mathfrak{f}_{xw}}{\partial z} + \frac{\partial \mathfrak{f}_{wz}}{\partial x} + \frac{\partial \mathfrak{f}_{zx}}{\partial w} &= 0, \\
\frac{\partial \mathfrak{f}_{yw}}{\partial x} + \frac{\partial \mathfrak{f}_{wx}}{\partial y} + \frac{\partial \mathfrak{f}_{xy}}{\partial w} &= 0, \\
\frac{\partial \mathfrak{f}_{uw}}{\partial x} + \frac{\partial \mathfrak{f}_{wx}}{\partial u} + \frac{\partial \mathfrak{f}_{xu}}{\partial w} &= 0, \\
\frac{\partial \mathfrak{f}_{uw}}{\partial y} + \frac{\partial \mathfrak{f}_{wy}}{\partial u} + \frac{\partial \mathfrak{f}_{yu}}{\partial w} &= 0, \\
\frac{\partial \mathfrak{f}_{uw}}{\partial z} + \frac{\partial \mathfrak{f}_{wz}}{\partial u} + \frac{\partial \mathfrak{f}_{zu}}{\partial w} &= 0.
\end{aligned}\right\} \qquad \text{(II)}$$

---

[1]) G. Mie, Ann. d. Phys. **40**, 25, 1913.

[2]) G. Nordström, diese Zeitschr. **13**, 1126, 1912; Ann. d. Phys. **40**, 872, 1913; **42**, 533, 1913.

[3]) H. Minkowski, Gött. Nachr. 1908, S. 58.

Die Gleichungssysteme sind beide vollkommen symmetrisch in bezug auf x, y, z, u, w. Sie haben natürlich zunächst keine physikalische Bedeutung; setzt man aber alle partiellen Ableitungen nach w gleich null, so findet man, daß sie in die Feldgleichungen des elektromagnetischen Feldes und des Gravitationsfeldes übergehen, indem $\mathfrak{s}_x$, $\mathfrak{s}_y$, $\mathfrak{s}_z$, $\mathfrak{s}_u$ die Komponenten des Viererstroms sind und $-\frac{1}{c}\mathfrak{s}_w$ die Ruhdichte der gravitierenden Masse[1] ist. Die vier ersten Gleichungen in den beiden Systemen sind nunmehr die Maxwellschen Gleichungen in der von Minkowski gegebenen Form; die letzte Gleichung (I) ist die Fundamentalgleichung der Gravitation, und die sechs übrigen Gleichungen (II) drücken die Wirbellosigkeit des Gravitationsvektors aus.

Diese Deutung der Gleichungen (I), (II) zeigt, daß es berechtigt ist, die vierdimensionale Raumzeitwelt als eine durch eine fünfdimensionale Welt gelegte Fläche aufzufassen. In jener fünfdimensionalen Welt sind die $\mathfrak{s}_m$ Komponenten eines Fünfervektors und die $\mathfrak{f}_{mn}$ Komponenten eines Zehnervektors; der letztere Vektor charakterisiert vollständig den physikalischen Zustand des Äthers. Die fünfdimensionale Welt hat eine ausgezeichnete Achse, die w-Achse; die vierdimensionale Raumzeitwelt steht senkrecht zu dieser Achse und in allen ihren Punkten sind die Ableitungen sämtlicher Komponenten von $\mathfrak{f}$ nach w gleich null.

Die Komponenten von $\mathfrak{f}$ lassen sich durch ein Fünferpotential mit den Komponenten Φ_x, Φ_y, Φ_z, Φ_u, Φ_w ausdrücken, indem für jede Komponente von $\mathfrak{f}$

$$\mathfrak{f}_{mn} = \frac{\partial \Phi_n}{\partial m} - \frac{\partial \Phi_m}{\partial n}. \tag{2}$$

Durch Differentiation der Gleichungen (I) findet man für den „Fünferstrom“ $\mathfrak{s}$ die Beziehung

$$\frac{\partial \mathfrak{s}_x}{\partial x} + \frac{\partial \mathfrak{s}_y}{\partial y} + \frac{\partial \mathfrak{s}_z}{\partial z} + \frac{\partial \mathfrak{s}_u}{\partial u} + \frac{\partial \mathfrak{s}_w}{\partial w} = 0, \tag{3}$$

und deswegen kann man für das Fünferpotential folgende sechs partielle Differentialgleichungen vorschreiben:

$$\frac{\partial \Phi_x}{\partial x} + \frac{\partial \Phi_y}{\partial y} + \frac{\partial \Phi_z}{\partial z} + \frac{\partial \Phi_u}{\partial u} + \frac{\partial \Phi_w}{\partial w} = 0, \tag{4}$$

$$\left.\begin{aligned} &\frac{\partial^2 \Phi_x}{\partial x^2} + \frac{\partial^2 \Phi_x}{\partial y^2} + \frac{\partial^2 \Phi_x}{\partial z^2} + \frac{\partial^2 \Phi_x}{\partial u^2} + \frac{\partial^2 \Phi_x}{\partial w^2} = \\ &\qquad = -\frac{1}{c}\mathfrak{s}_x, \\ &\frac{\partial^2 \Phi_w}{\partial x^2} + \frac{\partial^2 \Phi_w}{\partial y^2} + \frac{\partial^2 \Phi_w}{\partial z^2} + \frac{\partial^2 \Phi_w}{\partial u^2} + \frac{\partial^3 \Phi_w}{\partial w^2} = \\ &\qquad = -\frac{1}{c}\mathfrak{s}_w. \end{aligned}\right| \tag{5}$$

Von den fünf Gleichungen in (5) sind wegen Raumersparnis nur die erste und letzte ausgeschrieben. Wenn diese sechs Bedingungen erfüllt sind, befriedigen die Ausdrücke (2) identisch die Feldgleichungen (I), (II).

Unsere Formeln (I), (II) ermöglichen auch, den Impuls-Energiesatz für das kombinierte elektromagnetische und Gravitationsfeld in einheitlicher Weise aufzustellen. Um diesen Satz für die x-Richtung zu erhalten, multipliziere man diejenigen vier Gleichungen (I), die sich auf die übrigen Achsenrichtungen beziehen, mit $\mathfrak{f}_{xy}$, $\mathfrak{f}_{xz}$, $\mathfrak{f}_{xu}$, $\mathfrak{f}_{xw}$, und diejenigen sechs Gleichungen (II), die x enthalten, mit $\mathfrak{f}_{mn}$, wo m, n die beiden übrigen in der betreffenden Gleichung vorkommenden Indizes in richtiger Reihenfolge bedeuten. Die so erhaltenen zehn Gleichungen sind zu addieren, wonach eine einfache Umformung den gewünschten Satz gibt.

Wir wollen den Satz für die u-Richtung — also den Energiesatz — aufstellen und haben dann die drei ersten Gleichungen (I) mit $\mathfrak{f}_{ux}$, $\mathfrak{f}_{uy}$, $\mathfrak{f}_{uz}$, die letzte mit $\mathfrak{f}_{uw}$ zu multiplizieren. Weiter sind von den Gleichungen (II) die zweite, dritte, vierte, achte, neunte und zehnte mit bzw. $\mathfrak{f}_{yz}$, $\mathfrak{f}_{zx}$, $\mathfrak{f}_{xy}$, $\mathfrak{f}_{wx}$, $\mathfrak{f}_{wy}$, $\mathfrak{f}_{wz}$ zu multiplizieren. Wir erhalten durch Addition und nach einiger Umstellung der Terme

---

[1] Hierunter verstehe ich die Größe, die ich in den zitierten Aufsätzen mit $g \cdot \nu$ bezeichnet habe; es ist also

$$-\frac{1}{c}\mathfrak{s}_w = g \cdot \nu.$$

Man könnte ja allgemeiner für die Komponenten des Gravitationsvektors die Bezeichnungen $a\mathfrak{f}_{wx}$, $a\mathfrak{f}_{wy}$, $a\mathfrak{f}_{wz}$, $a\mathfrak{f}_{wu}$ einführen, wo a eine beliebige reelle oder imaginäre Konstante sein würde; dann würde man $-\frac{a}{c}\mathfrak{s}_w = g\nu$ haben. Es wird sich aber bei der Aufstellung des Impuls-Energiesatzes zeigen, daß a gleich $+1$ oder -1 sein muß, damit die letzte Gleichung (I) sich wirklich wie die übrigen verhalte.

$$\begin{aligned}
&\mathfrak{f}_{ux}\left(\frac{\partial \mathfrak{f}_{xy}}{\partial y}+\frac{\partial \mathfrak{f}_{xz}}{\partial z}+\frac{\partial \mathfrak{f}_{xw}}{\partial w}\right)+\\
&+\mathfrak{f}_{uy}\left(\frac{\partial \mathfrak{f}_{yx}}{\partial x}+\frac{\partial \mathfrak{f}_{yz}}{\partial z}+\frac{\partial \mathfrak{f}_{yw}}{\partial w}\right)+\\
&+\mathfrak{f}_{uz}\left(\frac{\partial \mathfrak{f}_{zx}}{\partial x}+\frac{\partial \mathfrak{f}_{zy}}{\partial y}+\frac{\partial \mathfrak{f}_{zw}}{\partial w}\right)+\\
&+\mathfrak{f}_{uw}\left(\frac{\partial \mathfrak{f}_{wx}}{\partial x}+\frac{\partial \mathfrak{f}_{wy}}{\partial y}+\frac{\partial \mathfrak{f}_{wz}}{\partial z}\right)+\\
&+\mathfrak{f}_{yz}\left(\frac{\partial \mathfrak{f}_{zu}}{\partial y}+\frac{\partial \mathfrak{f}_{uy}}{\partial z}\right)+\mathfrak{f}_{zx}\left(\frac{\partial \mathfrak{f}_{xu}}{\partial z}+\frac{\partial \mathfrak{f}_{uz}}{\partial x}\right)+\\
&+\mathfrak{f}_{xy}\left(\frac{\partial \mathfrak{f}_{yu}}{\partial x}+\frac{\partial \mathfrak{f}_{ux}}{\partial y}\right)+\mathfrak{f}_{wx}\left(\frac{\partial \mathfrak{f}_{uw}}{\partial x}+\frac{\partial \mathfrak{f}_{xu}}{\partial w}\right)+\\
&+\mathfrak{f}_{wy}\left(\frac{\partial \mathfrak{f}_{uw}}{\partial y}+\frac{\partial \mathfrak{f}_{yu}}{\partial w}\right)+\mathfrak{f}_{wz}\left(\frac{\partial \mathfrak{f}_{uw}}{\partial z}+\frac{\partial \mathfrak{f}_{zu}}{\partial w}\right)+\\
&+\mathfrak{f}_{ux}\frac{\partial \mathfrak{f}_{xu}}{\partial u}+\mathfrak{f}_{uy}\frac{\partial \mathfrak{f}_{yu}}{\partial u}+\mathfrak{f}_{uz}\frac{\partial \mathfrak{f}_{zu}}{\partial u}+\mathfrak{f}_{uw}\frac{\partial \mathfrak{f}_{wu}}{\partial u}+\\
&+\mathfrak{f}_{yz}\frac{\partial \mathfrak{f}_{yz}}{\partial u}+\mathfrak{f}_{zx}\frac{\partial \mathfrak{f}_{zx}}{\partial u}+\mathfrak{f}_{xy}\frac{\partial \mathfrak{f}_{xy}}{\partial u}+\\
&+\mathfrak{f}_{wx}\frac{\partial \mathfrak{f}_{wx}}{\partial u}+\mathfrak{f}_{wy}\frac{\partial \mathfrak{f}_{wy}}{\partial u}+\mathfrak{f}_{wz}\frac{\partial \mathfrak{f}_{wz}}{\partial u}=\\
&=\frac{1}{c}(\mathfrak{f}_{ux}\mathfrak{k}_x+\mathfrak{f}_{uy}\mathfrak{k}_y+\mathfrak{f}_{uz}\mathfrak{k}_z+\mathfrak{f}_{uw}\mathfrak{k}_w).
\end{aligned}$$

Eine einfache Umformung gibt die gesuchte Gleichung

$$\left.\begin{aligned}
&\frac{\partial}{\partial x}(\mathfrak{f}_{uy}\mathfrak{f}_{yx}+\mathfrak{f}_{uz}\mathfrak{f}_{zx}+\mathfrak{f}_{uw}\mathfrak{f}_{wx})+\\
&+\frac{\partial}{\partial y}(\mathfrak{f}_{ux}\mathfrak{f}_{xy}++)+\frac{\partial}{\partial z}(\mathfrak{f}_{ux}\mathfrak{f}_{xz}++)+\\
&+\frac{\partial}{\partial w}(\mathfrak{f}_{ux}\mathfrak{f}_{xw}++)+\frac{1}{2}\frac{\partial}{\partial u}(-\mathfrak{f}_{ux}^2-\\
&-\mathfrak{f}_{uy}^2-\mathfrak{f}_{uz}^2-\mathfrak{f}_{uw}^2+\mathfrak{f}_{yz}^2+\mathfrak{f}_{zx}^2+\\
&+\mathfrak{f}_{xy}^2+\mathfrak{f}_{wx}^2+\mathfrak{f}_{wy}^2+\mathfrak{f}_{wz}^2)=\\
&=\frac{1}{c}(\mathfrak{f}_{ux}\mathfrak{k}_x+\mathfrak{f}_{uy}\mathfrak{k}_y+\mathfrak{f}_{uz}\mathfrak{k}_z+\mathfrak{f}_{uw}\mathfrak{k}_w).
\end{aligned}\right\} \quad (6)$$

Die Größen in den Klammern links sind Komponenten eines fünfdimensionalen Tensors, und die Gleichung hat demnach die Form

$$\frac{\partial P_{ux}}{\partial x}+\frac{\partial P_{uy}}{\partial y}+\frac{\partial P_{uz}}{\partial z}+\frac{\partial P_{uu}}{\partial u}+\frac{\partial P_{uw}}{\partial w}=\mathfrak{K}_u. \quad (6\,\mathrm{a})$$

Indem man sie mit ic multipliziert und $\frac{\partial P_{uw}}{\partial w}$ gleich null setzt, erhält man die Energiegleichung in ihrer gewöhnlichen Form. Man findet so für die x-Komponente des Energiestromes

$$\left.\begin{aligned}
\mathfrak{S}_x &= ic(\mathfrak{f}_{uy}\mathfrak{f}_{yx}+\mathfrak{f}_{uz}\mathfrak{f}_{zx}+\mathfrak{f}_{uw}\mathfrak{f}_{wx})\\
&= c(\mathfrak{E}_y\mathfrak{H}_z-\mathfrak{E}_z\mathfrak{H}_y)-\frac{\partial \Phi_w}{\partial t}\frac{\partial \Phi_w}{\partial x}.
\end{aligned}\right\} \quad (7)$$

Der letzte Ausdruck ist mittels der Formeln (1) und (2) erhalten, und es sind die auftretenden Ableitungen nach w gleich null gesetzt. In vektoranalytischer Schreibweise lautet der Ausdruck für $\mathfrak{S}$

$$\mathfrak{S}=c[\mathfrak{E}\mathfrak{H}]-\frac{\partial \Phi_w}{\partial t}\nabla \Phi_w. \quad (7\,\mathrm{a})$$

Für die Energiedichte bekommt man ebenfalls aus (6) einen Ausdruck, der, vektoranalytisch geschrieben, lautet:

$$\psi=\frac{1}{2}\left\{\mathfrak{E}^2+\mathfrak{H}^2+(\nabla \Phi_w)^2+\frac{1}{c^2}\left(\frac{\partial \Phi_w}{\partial t}\right)^2\right\}. \quad (8)$$

Diese Ausdrücke für Energiestrom und Energiedichte setzen sich aus den bekannten Ausdrücken für die betreffenden Größen im elektromagnetischen Felde und im Gravitationsfelde additiv zusammen; was ja das gewünschte Resultat unserer Betrachtungen ist.

Man sieht leicht, daß die Behauptung in der Note S. 504 r. Sp. richtig ist, und daß also von den Komponenten des Zehnervektors $\mathfrak{f}$ und des Fünfervektors $\mathfrak{k}$ nur die mit einem Index u versehenen imaginär sind.

Durch Permutation der Indizes in Gleichung (6) erhält man vier andere Gleichungen. Die drei, die sich auf räumliche Achsenrichtungen beziehen, drücken natürlich den Impulssatz in bekannter Weise aus. Die Gleichung für die w-Richtung lautet nach Einführung der Feldstärken und des Gravitationspotentials Φ_w vektoranalytisch geschrieben

$$\begin{aligned}
&-\operatorname{div}\left\{[\mathfrak{H},\nabla\Phi_w]+\frac{1}{c}\mathfrak{E}\frac{\partial \Phi_w}{\partial t}\right\}+\frac{1}{c}\frac{\partial}{\partial t}(\mathfrak{E}\nabla\Phi_w)+\\
&+\frac{1}{2}\frac{\partial}{\partial w}\left\{\mathfrak{H}^2-\mathfrak{E}^2-(\nabla\Phi_w)^2+\frac{1}{c^2}\left(\frac{\partial \Phi_w}{\partial t}\right)^2\right\}=\\
&=-\frac{1}{c}\left\{\mathfrak{k}\nabla\Phi_w+\mathfrak{k}_u\frac{\partial \Phi_w}{\partial u}\right\}.
\end{aligned}$$

Es ist vorläufig nicht einzusehen, daß dieser Gleichung eine physikalische Bedeutung zukäme.

Die obige Betrachtungsweise bietet, wie wir gesehen haben, formale Vorteile, indem nach ihr das elektromagnetische Feld und das Gravitationsfeld als ein einziges Feld sich auffassen lassen. Ein neuer physikalischer Inhalt ist natürlich den Gleichungen damit nicht gegeben. Ich halte es indessen nicht für ausgeschlossen, daß die gefundene formale Symmetrie einen tieferen Grund haben könnte. Auf die Möglichkeiten, die man sich in dieser Hinsicht denken könnte, will ich jedoch hier nicht eingehen.

Zusammenfassung. Es wird gezeigt, daß eine einheitliche Behandlung des elektromagnetischen Feldes und des Gravitationsfeldes möglich ist, wenn man die vierdimensionale Raumzeitwelt als eine durch eine fünfdimensionale Welt gelegte Fläche auffaßt.

Helsingfors, 30. März 1914.

(Eingegangen 3. April 1914.)

ON THE UNITY PROBLEM OF PHYSICS*

By TH. KALUZA
in Königsberg

(Presented by Mr. Einstein on 8 December 1921)

Alongside the metric tensor of the 4-dimensional manifold (interpreted as a tensor potential of gravity) a general relativistic description of world-phenomena requires also an electromagnetic four potential A_μ.**

The remaining dualism of gravity and electricity, while not lessening the theory's enthralling beauty, nevertheless sets the challenge of overcoming it through a fully unified world view.

A few years ago H. Weyl[1] made a surprisingly bold thrust towards the solution of this problem, one of the great favorite ideas of the human spirit. Through a new radical revision of the geometric foundations, he obtains along with the tensor $g_{\mu\nu}$ a kind of fundamental metric *vector* which he interprets as the electromagnetic potential A_μ. The complete world metric then becomes the common source of all natural events.

The same will be accomplished here by different means.

Disregarding the difficulties which accompany H. Weyl's deepseated theory, ideally, one can imagine an even more perfect realization of the quest for unity: the gravitational and electromagnetic fields stem from a single universal tensor. I wish to show here, that such a tight union of both world powers seems possible in principle.

---

* Translator's note: The original paper in German appeared in Sitzungsberichte der K. Preussischen Akademie der Wissenschaften zu Berlin on p. 966 (1921). Copyright: Akademie der Wissenschaften der DDR.

** Translator's note: We changed the original notation for the electromagnetic four-potential which was q_μ.

[1] Sitzungsber. d. Berl. Akad. 1918 p 465.

The fact that the electromagnetic field components $F_{\kappa\lambda}$ are given by a curl, and even more suggestively the unmistakable correspondence in the buildup of gravitational and electromagnetic equations,[2] formally incite the conjecture that the quantities $\frac{1}{2}F_{\kappa\lambda} = \frac{1}{2}(A_{\kappa,\lambda} - A_{\lambda,\kappa})$ could somehow be truncated three index symbols $\begin{bmatrix} \iota\lambda \\ \kappa \end{bmatrix} = \frac{1}{2}(g_{\iota\kappa,\lambda} + g_{\kappa\lambda,\iota} - g_{\iota\lambda,\kappa})$. Giving in to this thought, one is pushed with a great degree of finality onto a not very inviting path: there being no further three index symbols in a four dimensional world, beyond those already used up in gravity, the above interpretation of $F_{\kappa\lambda}$ can hardly be maintained short of making the strongly alienating decision of calling a fifth dimension to the rescue.

Although all our previous physical experience hardly provides any suggestion of such an extra world-parameter, we are certainly free to consider our space-time to be a four-dimensional part of an R_5; one then has to take into account the fact that we are only aware of the space-time variation of quantities, by making their derivatives with respect to the new parameter vanish or by considering them to be small as they are of higher order ("cylinder condition"). The fear that by this condition the introduction of the fifth dimension would be revoked is unwarranted, because of the linkage of the world-parameters in the three-index-symbols.

* * *

We therefore enter R_5 and extend Einstein's Ansatz to it. Here, in addition to ordinary variables x^1 to x^4, x^5 comes into play as a new parameter.* If the metric tensor of this R_5 is expressed as g_{rs},** the three-index-symbols $\begin{bmatrix} lm \\ n \end{bmatrix}$

---

[2] See H. Thirring, Phys. Ztschr. **19**, p. 204.

* Translator's note: Kaluza uses x^0 instead of x^5.

** Latin indices always run from 1 to 5 and Greek ones only from 1 to 4.

denoted here by $-\Gamma_{lmn}$ become, by virtue of the cylinder condition,

$$2\Gamma_{\lambda\mu\nu} = g_{\lambda\mu,\nu} - g_{\mu\nu,\lambda} - g_{\nu\lambda,\mu} \text{ (as before)}$$

$$2\Gamma_{5\mu\nu} = g_{5\mu,\nu} - g_{5\nu,\mu} ,$$

$$2\Gamma_{\lambda\mu5} = -g_{5\lambda,\mu} - g_{5\mu,\lambda} ,$$

$$2\Gamma_{55\nu} = g_{55,\nu}, \; 2\Gamma_{5\mu5} = -g_{55,\mu}, \; 2\Gamma_{555} = 0 \tag{1}$$

At first sight this result is hardly encouraging. Indeed $\Gamma_{5\mu\nu}$ appear in the form of curls, but the ten $\Gamma_{\lambda\mu5}$, which would have to be of electric nature according to our interpretation, are in danger of obstructing the way. Nevertheless we continue and, in order to keep $\Gamma_{5\mu\nu}$ proportional to $F_{\mu\nu}$ we set*

$$g_{5\mu} = 2\alpha A_\mu, \; g_{55} = 2\phi , \tag{2}$$

so that the metric tensor of R_5 substantially becomes the gravitational tensor potential framed by the electromagnetic four potential: The role of the component ϕ in the corner remains unsettled for the time being. Denoting by $\Sigma_{\mu\nu}$ the sums $A_{\mu,\nu} + A_{\nu,\mu}$, the counterparts of the $F_{\mu\nu}$, we have

$$\Gamma_{5\mu\nu} = \alpha F_{\mu\nu}, \; \Gamma_{\mu\nu5} = -\alpha\Sigma_{\mu\nu} ,$$

$$\Gamma_{55\mu} = -\Gamma_{5\mu5} = \phi_{,\mu} . \tag{3}$$

Consequently the thirty-five new three-index-symbols (five of which vanish) comprise the electromagnetic field $F_{\mu\nu}$, its "associated" field $\Sigma_{\mu\nu}$ and the gradient[3] of ϕ. Moreover, the general equation

$$(\Gamma_{lmn}+\Gamma_{mnl}+\Gamma_{nlm})_{,k} = \Gamma_{klm,n}+\Gamma_{kmn,l}+\Gamma_{knl,m} \tag{4}$$

yields, by virtue of the cylinder condition, the well-known relations

---

[*] Translator's note: Kaluza's notation for ϕ was g.

[3] In the four-dimensional sense.

$$F_{\mu\nu,\lambda}+F_{\nu\lambda,\mu}+F_{\lambda\mu,\nu}=0$$

$$\phi_{,\mu\nu}=\phi_{,\nu\mu}\ . \tag{4a}$$

* * *

We now restrict, as usual, the choice of parameters by $g = \det g_{rs} = -1$ and let g_{rs} differ only a little from the "Euclidean" value $-\delta_{rs}$ (approximation I). With $\Gamma^{n}_{lm} = -\left\{ {lm \atop n} \right\} = -\Gamma_{lmn}$, the components of the second four-index-tensor[**] of interest here become

$$\{\mu\nu,\lambda 5\} = \alpha F^{\nu}_{\mu,\lambda}\ ,\ \ \{\mu 5,5\nu\} = -\phi_{,\mu\nu}\ ,$$

$$\{\mu\nu,55\} = \{\mu 5,55\} = \{55,55\} = 0\ . \tag{5}$$

Fortunately the associated field in Eq. (3) does not participate here. Of the electric quantities only the fields $F_{\mu\nu}$ play a role in determining the curvature of R_5. By further constructing the contracted tensor[*] $R_{mn} = \{ml,ln\}$, we find according to our assumptions (in the well-known notation),[**]

$$R_{\mu\nu} = \Gamma^{\lambda}_{\mu\nu,\lambda}\ ,\ \ \text{(as before)},$$

$$R_{5\nu} = -\alpha\partial^{\mu}F_{\mu\nu}\ ,$$

$$R_{55} = - \quad \phi\ . \tag{6}$$

Therefore the fifteen components of the curvature tensor[†] are grouped into the left-hand-sides of 1) the ordinary field equations of gravitation; 2) the basic electromagnetic equations, and 3) a Poisson equation for the yet uninterpreted ϕ. Therein lies the foremost justification of our Ansatz and the hope to consider

---

[**] Translator's note: The symbol $\{\mu\lambda,\nu\rho\}$ corresponds to the Riemann-Christoffel curvature tensor $R^{\lambda}_{\mu\nu\rho}$.

[*] Translator's note: Ricci tensor.

[**] Translator's note: Originally the right-hand-side of the second equation of Eq. (6) was written as $-\alpha\Delta\iota\nu_{\mu}F$.

[†] Translator's note: Ricci tensor.

gravitation and electricity as manifestations of a universal field.

* * *

In R_5 the energy-momentum tensor of the matter which governs the right-hand-side of the field equations, is given in the approximation I by

$$T_{mn} = T^{mn} = \mu_0 u^m u^n \; . \tag{7}$$

$$(\mu_0 = \text{rest mass density}, \; \mu^n = dx^n/ds, ds^2 = g_{mn} dx^m dx^n).$$

Since now $R_{5\mu} = -\kappa T_{5\mu}$ (for all three kinds of equations), from the Maxwell equations according to Eq. (6) yield for the components of the four-current,

$$I^\mu = \rho_0 v^\mu = (\kappa/\alpha) T_{5\mu} = (\kappa/\alpha)\mu_0 u^5 u^\mu \; . \tag{8}$$

$$(\rho_0 = \text{rest charge density}, \; v^\mu = dx^\mu/d\sigma, \; d\sigma^2 = g_{\mu\nu} dx^\mu dx^\nu) \; .$$

In other words the space-time energy-momentum tensor is essentially framed by the current density.

We further pursue the investigation under the assumption $u^1, u^2, u^3, u^5 << 1, u^4 \sim 1$ (approximation II). This approximation implies not only small velocity but also a very small specific charge ρ_0/μ_0 of the moving matter; in fact, as then $d\sigma^2 \sim ds^2$ and $v^\mu \sim u^\mu$, it follows from Eq. (8), if one sets* $\alpha = \sqrt{\kappa/2} = 3.06 \times 10^{-14}$ by anticipation,[4] that

$$\rho_0 = (\kappa/\alpha)\mu_0 u^5 = 2\alpha\mu_0 u^5 << \mu_0 \; . \tag{8a}$$

First of all, this equation teaches us again that in this case we have to understand the electric charge essentially as the fifth component of the energy-momentum of the matter "moving obliquely" to the spaces $x^5 = $ const.. A further fusion of two formerly heterogeneous basic concepts appears to be achieved thereby.

---

* **Translator's note:**
$\kappa = 8\pi G/c^2 = 1.86 \times 10^{-27} cm/gr, \; \sqrt{\kappa/2} = 3.06 \times 10^{-14} (cm/gr)^{1/2}$

[4] For the sake of the equation of motion; see the next section.

Because finally T_{11}, T_{22}, T_{33}, $T_{55} \sim 0$ in approximation II, we find according to Eq. (7)

$$T \equiv g^{mn} T_{mn} = -T_{44} = -\mu_0 \, , \tag{9}$$

so that for the ordinary form of the field equation of the first kind we have

$$R_{55} = -R_{44} = \kappa \mu_0 / 2 \, . \tag{10}$$

The corner potential ϕ is thus found, due to Eq. (6), to be essentially the negative of the gravitational potential, while $= g_{44}/2$ keeps its former meaning.

* * *

After having thus disposed of the standard quantities in the field equations in a satisfactory way, we are still faced with the question whether the "geodesic" equation of motion in R_5,

$$\dot{u}^l \equiv du^l / ds = \Gamma^l_{mn} u^m u^n \, , \tag{11}$$

still represents the motion of charged matter in the gravitational and electromagnetic field in accordance with experiments. In approximation II this is immediately the case. Due to the interchangeability of ds and $d\sigma$ one obtains according to Eq. (3)

$$\overrightarrow{v}^{\lambda} \equiv dv^{\lambda} / d\sigma = \Gamma^{\lambda}_{\mu\nu} v^{\mu} v^{\nu} + 2\alpha F^{\lambda}_{\mu} u^5 v^{\mu} - \phi' \ {}^{\lambda} (u^5)^2 \, , \tag{11a}$$

i.e., due to the smallness of the term with $(u^5)^2$, for the force density one finds

$$\pi^{\lambda} \equiv \mu_0 \overrightarrow{v}^{\lambda} = \Gamma^{\lambda}_{\mu\nu} T^{\mu\nu} + F^{\lambda}_{\mu} I^{\mu} \ (\alpha = \sqrt{\kappa/2}, \text{ see the footnote 5}). \tag{12}$$

The total force thus automatically splits into a gravitational and an electromagnetic part of the usual form.

For the fifth component of Eq. (11) there remains in the end only

$$\dot{u}^5 = \alpha \Sigma_{44} = 2\alpha A_{4,4} \, , \tag{11b}$$

---

[5] See equation (8a).

so that, in the quasi-statics characterized by approximation II, $(d/dx^4)(\rho_0/\mu_0) = 2\kappa A_{4,4}$ is of higher order.[5] The necessary constancy of ρ_0 seems to be guaranteed accordingly.

Hence also for the equations of motion the associated field remains insignificant in our approximation.

* * *

If approximation II would be in accordance with reality, the desired unitarian theory would have essentially been realized above in a satisfactory manner. A unique potential tensor generates a universal field which splits, under ordinary conditions, into a gravitational and an electromagnetic part.

Now, however, the matter at least in its smallest building blocks is not at all weakly charged: following Weyl's expression, its "unrest in the small" is contrasted with its "rest in the large", and this holds, in the above mentioned view, quite specifically for the new world-parameter x^5. For the electron and nucleus, ρ_0/μ_0 is not small and so the "velocity" component u^5 is not at all small. In the form constrained by approximation II the theory can at best coarsely describe macroscopic phenomena, and a fundamental problem arises concerning its very applicability to those elementary particles.

If one now, nevertheless, tries to describe the motion of the electron by geodesics in R_5, one immediately meets with a serious difficulty[6] which threatens to demolish what we have built. The difficulty lies simply in the fact that, with the rigid application of the previous assumptions, u^5 for the electrons becomes enormously large owing to $e/m = 1.77 \times 10^7$ (in units of the light seconds) so that the last term in Eq. (11a), instead of vanishing, takes a value which surpasses everything and mocks all experience, even if everything else remains formally as before. Now the larger value of u^5 anyhow implies modifications of the theory (the interchangeability of ds and $d\sigma$ is lost), but it hardly seems possible to develop the theory in the old framework alone, without any new hypotheses.

---

[6] I owe notice of the following inconsistencies to Mr. Einstein's valuable interest in the birth of this work.

On the other hand I think -- with due reservations -- that there is a way, which, if it achieves its goal, could lead to an even more satisfying point of view. Since for velocities of the material source of the field which are not too large, $R_{55} \sim R_{44}$ even for arbitrary u_5, the two gravitation-like terms in (11a) appear, under an appropriate interpretation of the so far irrelevant character of x^5, with opposite signs. It thus appears that, surrendering the in any case questionable gravitational constant κ, the two contradicting orders of magnitude could be reconciled, and gravitation would become a kind of difference effect. This line of thought is enticing due to the prospect of being able to assign to that constant the role of a statistical quantity. At present the consequences of this hypothesis are not clear; furthermore there are other possibilities to be investigated. In any case, every Ansatz which claims universal validity is threatened by the sphinx of modern physics, quantum theory.

* * *

Even in the face of all the physical and epistemological difficulties which we have seen piling up against the conception presented here, it is still hard to believe that all these relations in their virtually unsurpassed formal unity, should amount to the mere alluring play of a capricious accident. Should more than an empty formalism be found to reside behind these presumed connections, we would then face a new triumph of Einstein's general relativity, whose appropriate application to a five-dimensional world is our main concern here.

This is a modified version (by P.G.O.F.) of earlier translations by T. Muta and C. Hoenselaers.

Zum Unitätsproblem der Physik.

Von Th. Kaluza
in Königsberg.

(Vorgelegt von Hrn. Einstein am 8. Dezember 1921 [s. oben S. 859].)

In der allgemeinen Relativitätstheorie muß zur Charakterisierung des Weltgeschehens dem als Tensorpotential der Gravitation gedeuteten metrischen Fundamentaltensor der vierdimensionalen Weltmannigfaltigkeit $g_{\mu\nu}$ noch das elektromagnetische Viererpotential q_{μ} zur Seite treten.

Der so auch hier verbleibende Dualismus von Gravitation und Elektrizität nimmt zwar jener Theorie nichts von ihrer bestrickenden Schönheit, fordert aber aufs neue zu seiner Überwindung durch ein restlos unitarisches Weltbild heraus.

Einen überraschend kühnen Vorstoß zur Lösung dieses Problems, das zu den großen Lieblingsideen des Menschengeistes gehört, hat vor einigen Jahren H. Weyl[1] unternommen, der bei einer nochmaligen, radikalen Revision der geometrischen Grundlegung neben dem Tensor $g_{\mu\nu}$ noch eine Art metrischen Fundamentalvektor erhält und als elektromagnetisches Potential q_{μ} interpretiert: Die vollständige Weltmetrik wird dort als gemeinsamer Quell alles Naturgeschehens hingestellt.

Auf einem anderen Wege soll hier das gleiche Ziel erstrebt werden.

Von den Schwierigkeiten abgesehen, die die Durchführung jener tiefgründigen Theorie H. Weyls begleiten, wäre auch ideell eine noch vollkommenere Verwirklichung des Unitätsgedankens vorstellbar: Gravitations- und elektromagnetisches Feld entspringen einem einzigen universellen Tensor. — Ich möchte nun zeigen, daß eine solch enge Union beider Weltmächte prinzipiell möglich erscheint.

* * *

Die Rotorform der elektromagnetischen Feldkomponenten $F_{\varkappa\lambda}$, noch mehr aber das unverkennbare formale Entsprechen im Bau der Gravitations- und der elektromagnetischen Gleichungen[2] fordern förmlich die Vermutung heraus, die $\frac{1}{2} F_{\varkappa\lambda} = \frac{1}{2}(q_{\varkappa\,.\,\lambda} - q_{\lambda\,.\,\varkappa})$[1] könnten irgendwie verstümmelte Dreizeigergrößen $\begin{bmatrix} i\lambda \\ \varkappa \end{bmatrix} = \frac{1}{2}\,(g_{i\varkappa\,.\,\lambda} + g_{\varkappa\lambda\,.\,i} - g_{i\lambda\,.\,\varkappa})$ sein. Gibt man diesem Gedanken Raum, so sieht man sich mit großer Bestimmtheit auf einen zunächst allerdings wenig einladenden Weg gedrängt: Da nämlich in einer vierdimensionalen Welt außer den bereits für die Feldkomponenten der Gravitation verbrauchten Dreizeigergrößen keine weiteren existieren, so läßt sich jene Auffassung der $F_{\varkappa\lambda}$ kaum anders halten, als daß man sich zu dem wohl stark befremdenden Entschluß aufrafft, eine neue, fünfte Weltdimension zu Hilfe zu rufen.

---

[1] Sitzungsber. d. Berl. Akad. 1918 p. 465.

[2] Vgl. dazu auch H. Thirring, Phys. Ztschr. 19 p. 204.

Nun birgt zwar unser bisheriger physikalischer Erfahrungsschatz sonst kaum einen Hinweis auf einen solchen überzähligen Weltparameter, doch steht es ja frei, unsere Raumzeitwelt als vierdimensionalen Anteil an einem R_5 anzusehen; nur muß man dann der Tatsache, daß wir niemals andere als raumzeitliche Änderungen von Zustandsgrößen bemerken, dadurch Rechnung tragen, daß man deren Ableitungen nach dem neuen Parameter gleich Null setzt bzw. als von höherer Ordnung klein behandelt (»Zylinderbedingung«). Die Befürchtung, es könnte dadurch die Einführung der fünften Dimension wieder rückgängig gemacht werden, ist wegen der Verkettung der Weltparameter in den Dreizeigergrößen unbegründet.

* * *

Wir begeben uns mithin in einen R_5 und übertragen auf ihn die Einsteinschen Ansätze; dabei mag x^0 als neuer Parameter neben die gewohnten x^1 bis x^4 treten. Bedeutet g_{rs}[2] den metrischen Fundamentaltensor dieses R_5, so werden kraft der Zylinderbedingung die hier mit $-\Gamma_{ikl}$ bezeichneten Dreizeigergrößen $\begin{bmatrix} ik \\ l \end{bmatrix}$:

$$\begin{aligned} &2\Gamma_{\varkappa\lambda\mu} = g_{\varkappa\lambda\,.\,\mu} - g_{\lambda\mu\,.\,\varkappa} - g_{\mu\varkappa\,.\,\lambda} \text{ (wie früher)}, \\ &2\Gamma_{0\varkappa\lambda} = g_{0\varkappa\,.\,\lambda} - g_{0\lambda\,.\,\varkappa}, \quad 2\Gamma_{\varkappa\lambda 0} = -(g_{0\varkappa\,.\,\lambda} + g_{0\lambda\,.\,\varkappa}), \\ &2\Gamma_{00\varkappa} = g_{00\,.\,\varkappa}, \quad 2\Gamma_{0\varkappa 0} = -g_{00\,.\,\varkappa}, \quad 2\Gamma_{000} = 0. \end{aligned} \tag{1}$$

Das Ergebnis ermutigt zunächst wenig: Zwar erscheinen die $\Gamma_{0\varkappa\lambda}$ wirklich in Rotorform, doch drohen die zehn $\Gamma_{\varkappa\lambda 0}$, die bei der angestrebten Deutung auch elektrischer Natur sein müßten, den Weg zu verlegen. Wir verfolgen ihn trotzdem und setzen, um die $\Gamma_{0\varkappa\lambda}$ den $F_{\varkappa\lambda}$ proportional zu erhalten:

---

[1] Durch einen Punkt abgesonderte Indizes sollen Differentiation nach den zugehörigen Weltparametern anzeigen.

[2] Lateinische Indizes sollen stets von 0, griechische nur von 1 bis 4 laufen.

$$g_{0\varkappa} = 2\alpha q_{\varkappa}, \quad g_{00} = 2\mathfrak{g}, \tag{2}$$

so daß der metrische Fundamentaltensor des R_5 im wesentlichen das mit dem elektromagnetischen Viererpotential geänderte Tensorpotential der Gravitation wird; die Rolle der Eckkomponente $\mathfrak{g}$ bleibt vorerst unbestimmt. Schreibt man noch für die den $F_{\varkappa\lambda}$ entsprechenden Summen $q_{\varkappa\,.\,\lambda} + q_{\lambda\,.\,\varkappa}$ kurz $\Sigma_{\varkappa\lambda}$, so hat man:

$$\Gamma_{0\varkappa\lambda} = \alpha F_{\varkappa\lambda}, \; \Gamma_{\varkappa\lambda 0} = -\alpha \Sigma_{\varkappa\lambda}, \; \Gamma_{00\varkappa} = -\Gamma_{0\varkappa 0} = \mathfrak{g}_{\,.\,\varkappa}. \tag{3}$$

Durch das elektromagnetische Feld $F_{\varkappa\lambda}$, sein »Nebenfeld« $\Sigma_{\varkappa\lambda}$ sowie den Gradienten[1] von $\mathfrak{g}$ sind somit die fünfunddreißig neuen Dreizeigergrößen (von denen fünf verschwinden) erschöpft. Dabei entstehen aus der umfassenden Gleichung

$$(\Gamma_{ikl} + \Gamma_{kli} + \Gamma_{lik})_{\,.\,m} = \Gamma_{mik\,.\,l} + \Gamma_{mkl\,.\,i} + \Gamma_{mli\,.\,k} \tag{4}$$

vermöge der Zylinderbedingung die bekannten Beziehungen:

$$F_{\varkappa\lambda\,.\,\mu} + F_{\lambda\mu\,.\,\varkappa} + F_{\mu\varkappa\,.\,\lambda} = 0 \text{ und } \mathfrak{g}_{\,.\,\varkappa\lambda} = \mathfrak{g}_{\,.\,\lambda\varkappa}. \tag{4a}$$

* * *

Wir beschränken jetzt, wie üblich, die Parameterwahl durch $g = |g_{rs}| = -1$ und lassen (Näherung I) die g_{rs} nur wenig von den »euklidischen« Werten $-\delta_{rs}$ abweichen. Mit $\Gamma_{ik}^{l} = -\left\{ \begin{matrix} ik \\ l \end{matrix} \right\} = -\Gamma_{ikl}$ werden dann die hier interessierenden Komponenten des zweiten Vierzeigertensors:

$$\begin{gathered} \{\varkappa\lambda, \mu 0\} = \alpha F^{\lambda}_{\varkappa\,.\,\mu}, \; \{\varkappa 0, 0\lambda\} = -\mathfrak{g}_{\,.\,\varkappa\lambda}, \\ \{\varkappa\lambda, 00\} = \{\varkappa 0, 00\} = \{00, 00\} = 0. \end{gathered} \tag{5}$$

Hier tritt erfreulicherweise das Nebenfeld aus Gl. (3.) nicht mehr auf: Von elektrischen Größen sind es die Feldableitungen allein, die die Krümmung des R_5 mitbestimmen. Bildet man vollends den verjüngten Tensor $R_{ik} = \{ir, rk\}$, so gilt kraft unserer Annahmen (in bekannter Schreibart):

$$\begin{aligned} R_{\mu\nu} &= \Gamma^{\varrho}_{\mu\nu\,.\,\varrho} \text{ (wie früher)}, \\ R_{0\mu} &= -\alpha \Delta\iota\upsilon_{\mu} F, \\ R_{00} &= -\Box \mathfrak{g}. \end{aligned} \tag{6}$$

Die fünfzehn Komponenten des Krümmungstensors zerfallen somit in die linken Seiten 1. der alten Feldgleichungen der Gravitation, 2. der elektromagnetischen Grundgleichungen, 3. einer Poissonschen Gleichung für das noch ungedeutete g. Darin liegt eine erste Rechtfertigung für unsern Ansatz und für die Hoffnung, Gravitation und Elektrizität als Äußerungen eines universellen Feldes zu erkennen.

---

[1] Vierdimensional zurückgedeutet.

* * *

Für den die rechten Seiten der Feldgleichungen beherrschenden Energietensor der Materie gilt in R_5 in Näherung I:

$$T_{ik} = T^{ik} = \mu_0 u^i u^k \tag{7}$$

$$\left(\mu_0 = \text{Maßenruhdichte},\ u^r = \frac{dx^r}{ds},\ ds^2 = g_{lm}dx^l dx^m\right).$$

Da nun (für alle drei Arten der Feldgleichungen) $R_{0\mu} = -\varkappa T_{0\mu}$ ist, so fordern die Maxwellschen Gleichungen gemäß (6) für die Komponenten des Viererstroms:

$$\mathrm{I}^\mu = \rho_0 v^\mu = \frac{\varkappa}{\alpha} T_{0\mu} = \frac{\varkappa}{\alpha} \mu_0 u^0 u^\mu \tag{8}$$

$$\left(\rho_0 = \text{Ladungsruhdichte},\ v^\varrho = \frac{dx^\varrho}{d\sigma},\ d\sigma^2 = g_{\lambda\mu}dx^\lambda dx^\mu\right);$$

der raumzeitliche Energietensor ist also im wesentlichen mit der Stromdichte zu rändern.

Wir führen die weitere Untersuchung zunächst unter der Annahme: $u^0, u^1, u^2, u^3 \ll 1$, $u^4 \backsim 1$ (Näherung II). Dies bedingt neben kleiner Geschwindigkeit weiterhin sehr geringe spezifische Ladung $\frac{\rho_0}{\mu_0}$ der bewegten Materie; denn, weil dann $d\sigma^2 \backsim ds^2$, $v^\varrho \backsim u^\varrho$ wird, folgt aus (8), wenn man noch vorwegnehmend[1] $\alpha = \sqrt{\frac{\varkappa}{2}} = 3.06 \cdot 10^{-14}$ setzt:

$$\rho_0 = \frac{\varkappa}{\alpha} \mu_0 u^0 = 2\alpha\mu_0 u^0 \ll \mu_0 . \tag{8a}$$

Vor allem lehrt aber diese Gleichung, daß wir in diesem Falle die elektrische Ladung im wesentlichen als fünfte Impulskomponente der »schräg« zu den Räumen $x^0 = \text{const.}$ »bewegten« Masse verstehen dürfen: Eine weitere Verschmelzung zweier sonst heterogener Grundbegriffe erscheint damit vollzogen.

---

[1] Um der Bewegungsgleichungen willen; s. nächsten Abschnitt.

Da endlich in Näherung II T_{00}, T_{11}, T_{22}, $T_{33} \infty 0$ sind, wird nach (7):

$$T = g^{ik} T_{ik} = -T_{44} = -\mu_0 , \tag{9}$$

mithin für die gewohnte Form der Feldgleichungen 1. Art:

$$R_{00} = -R_{44} = \frac{\varkappa}{2} \mu_0 . \tag{10}$$

Das Eckpotential $\mathfrak{g}$ erweist sich also nach (6) hier im wesentlichen als negatives Gravitationspotential, während $\mathfrak{G} = \frac{g_{44}}{2}$ die alte Bedeutung behält.

* * *

Nachdem so über die maßgebenden Größen der Feldgleichungen in befriedigender Weise verfügt ist, spitzt sich die Frage darauf zu, ob die »geodätischen« Bewegungsgleichungen im R_5:

$$\dot{u}^l = \frac{du^l}{ds} = \Gamma^l_{rs} u^r u^s \tag{11}$$

nun auch die Bewegung geladener Materie im Gravitations- und elektromagnetischen Feld erfahrungstreu darstellen. In Näherung II ist dies ohne weiteres der Fall: Wegen der Vertauschbarkeit von ds und $d\sigma$ erhält man ja vermöge (3):

$$\bar{v}^\lambda = \frac{dv^\lambda}{d\sigma} = \Gamma^\lambda_{\varrho\sigma} v^\varrho v^\sigma + 2\alpha F^\lambda_\varkappa u^0 v^\varkappa - \mathfrak{g}_{\cdot\lambda} u^{0^2} , \tag{11 a}$$

d. h. wegen der Kleinheit des Gliedes mit u^{0^2} für die ponderomotorische Kraftdichte:

$$\pi^\lambda = \mu_0 \bar{v}^\lambda = \Gamma^\lambda_{\varrho\sigma} T^{\varrho\sigma} + F^\lambda_\varkappa \mathfrak{I}^\varkappa \left(\alpha = \sqrt{\frac{\varkappa}{2}} \text{ genommen; vgl. Anm.} \right). \tag{12}$$

Die Gesamtkraft spaltet sich also von selbst in einen gravitatorischen und einen elektromagnetischen Anteil der gewohnten Form.

Schließlich bleibt für die 0-Komponente von (11) nur:

$$\dot{u}^0 = \alpha \, \Sigma_{44} = 2\alpha q_{4\cdot 4} , \tag{11 b}$$

so daß in der durch Näherung II bedingten Quasistatik $\frac{d}{dx^4}\left(\frac{\rho_0}{\mu_0}\right) = 2\varkappa q_{4\cdot 4}$[1] klein von höherer Ordnung wird: Die erforderliche Konstanz von ρ_0 erscheint damit gewährleistet.

---

[1] Vgl. (8 a).

Auch für die Bewegungsgleichungen bleibt sonach das Nebenfeld in unserer Näherung bedeutungslos.

* * *

Entspräche Näherung II der Wirklichkeit, so wäre in Vorstehendem die erstrebte unitarische Theorie in ihren Hauptzügen befriedigend durchgeführt: Ein einziger Potentialtensor erzeugt ein universelles Feld, das sich unter gewöhnlichen Bedingungen in einen gravitatorischen und einen elektrischen Anteil spaltet.

Nun ist aber die Materie in ihren letzten Bausteinen zum mindesten durchaus nicht schwach geladen; ihrer »Ruhe im Großen« steht, mit H. Weyl zu reden, ihre »Unruhe im Kleinen« gegenüber, und dies gilt, in obiger Auffassung, ganz besonders für den neuen Weltparameter x^0: Beim Elektron oder dem H-Kern ist ja $\frac{\rho_0}{\mu_0}$ und damit die »Geschwindigkeits«-Komponente u^0 nichts weniger als klein! In der durch Näherung II bedingten Form kann also die Theorie höchstens grobphänomenologisch das makrophysische Geschehen zusammenfassen, und es schließt sich die Kernfrage nach ihrer Anwendbarkeit auf eben jene Urteilchen an.

Versucht man nun aber, die Bewegung des Elektrons durch eine Geodäte des R_5 zu beschreiben, so steht man sogleich vor einer ernsthaften Schwierigkeit[1], die den errichteten Bau einzustürzen droht. Sie besteht kurz darin, daß für das Elektron wegen $\frac{e}{m} = 1.77 \cdot 10^7$ (auf Lichtsekunde reduziert) u^0 bei starrer Übertragung der früheren Annahmen von so enorm hoher Größenordnung wird, daß das letzte Glied in (11a), anstatt zu verschwinden, einen alles überragenden und aller Erfahrung hohnsprechenden Wert annimmt, wenn auch sonst formal alles beim alten bleibt. Nun bedingt zwar der Übergang zu großen u^0 ohnehin Modifikationen (so entfällt die Ersetzbarkeit von ds durch $d\sigma$), doch will es kaum möglich erscheinen, die Theorie ohne neue Hypothese, allein im alten Rahmen durchzuführen.

---

[1] Einen Hinweis auf die folgende Unstimmigkeit verdanke ich dem wertvollen Interesse Hrn. Einsteins an dem Werden obiger Ansätze.

Dagegen glaube ich — mit allem Vorbehalt – in folgender Richtung einen Weg offen zu sehen, der, wenn er zum Ziele führt, einen wohl noch befriedigenderen Standpunkt erschlösse. Da bei nicht allzu großen Geschwindigkeiten der felderzeugenden Materie, auch für beliebige u^0, $R_{00} \backsim - R_{44}$ bleibt, nehmen die beiden gravitationsartigen Glieder in (11a) bei geeigneter Festsetzung des bisher noch völlig irrelevanten Realitätscharakters von x^0 entgegengesetztes Zeichen an, und es scheint dann unter Preisgabe der ohnedies etwas fragwürdigen Gravitationskonstante $\varkappa$ eine Versöhnung der widerstreitenden Größenordnungen zustande zu kommen, bei der die Gravitation als eine Art Differenzeffekt übrigbleibt. Dieser Weg besticht durch die Aussicht, jener Konstanten die Rolle einer statistischen Größe zuweisen zu können. Doch lassen sich zur Zeit die Konsequenzen dieser Hypothese noch zu wenig übersehen; auch sind noch anderweitige Möglichkeiten ins Auge zu fassen. Überhaupt droht ja jedem universelle Geltung heischenden Ansatz die Sphinx der modernen Physik, die Quantentheorie.

* * *

Trotz voller Würdigung der geschilderten physikalischen wie auch der erkenntnistheoretischen Schwierigkeiten, die sich vor der hier entwickelten Auffassung auftürmen, will es einem schwer werden, zu glauben, daß in all jenen an formaler Einheitlichkeit kaum zu überbietenden Beziehungen immer nur ein launischer Zufall sein lockendes Spiel treibt. Sollte es sich aber einmal bestätigen, daß mehr hinter den vermuteten Zusammenhängen steckt als nur ein leerer Formalismus, so würde dies entschieden einen neuen Triumph für EINSTEINS allgemeine Relativitätstheorie bedeuten, um deren sinngemäße Anwendung auf eine fünfdimensionale Welt es sich hier handelt.

Quantum theory and five dimensional theory of relativity

by

Oskar Klein
Kopenhagen

(received 28th April 1926)

On the following pages I should like to point out a simple connection between the theory for the relation of electromagnetism and gravitation proposed by Kaluza[1] on the one hand and the method given by de Broglie[2] and Schrödinger[3] for the treatment of quantum problems on the other. Kaluza's theory aims to link Einstein's ten gravitational potentials g_{ik} and the four electromagnetic potentials ϕ_i with the coefficients γ_{ik} of a line element of a Riemannian space which contains a fifth dimension in addition to the usual four. Also in electromagnetic fields the equations of motion of electric particles take the form of equations of geodesic lines. Interpreting these as equations of rays by viewing matter as a kind of wave propagation, one arrives almost naturally at a partial differential equation of second order which can be regarded as a generalization of the usual wave equation. Taking into account only those solutions of this equation which are periodic in the fifth dimension with a period connected to Planck's constant, one is led to the above mentioned quantum theoretical methods.

§1. Five dimensional theory of relativity. I shall begin by giving a short presentation of the five dimensional theory of relativity which is similar to but in some points distinct from Kaluza's theory.

Let us consider a five dimensional Riemannian line element for which we postulate a meaning independent of the system of coordinates. We write:

$$d\sigma = \sqrt{\Sigma\gamma_{ik}dx^i dx^k} \ , \qquad (1)$$

where the symbol Σ , as in what follows, indicates a summation from 0 to 4 over indices which appear twice. $x^0 \ldots x^4$ stand for the five coordinates of space. The 15 quantities γ_{ik} are the covariant components of a symmetric five dimensional tensor. To come from these to the quantities g_{ik} and ϕ_i of usual relativity, we have to make certain special assumptions. First, four of the coordinates, x^1, x^2, x^3, x^4, say, are to characterize the usual space-time. Second, the quantities γ_{ik} are not to depend on the fifth coordinate x^o. It follows that the permitted coordinate transformations are restricted to the following group[1] :

---

[1] Th. Kaluza, Sitzungsber. d. Berl. Akad. 1921, p. 966
[2] L. de Broglie, Ann. d. Phys. (10) 3, 22, 1925. Thèses, Paris. 1924.
[3] E. Schrödinger, Ann. d. Phys. 79, 361 and 489, 1926

$$\begin{aligned} x^o &= x^{o'} + \psi_o(x^{1'},x^{2'},x^{3'},x^{4'}), \\ x^i &= \psi_i(x^{1'},x^{2'},x^{3'},x^{4'}) \qquad (i = 1,2,3,4). \end{aligned} \tag{2}$$

We should actually have written constant times $x^{o'}$ instead of $x^{o'}$ in the first equation. Restricting this constant to the value unity is, however, quite irrelevant.

One can easily show that γ_{oo} remains invariant under the transformations (2). The assumption γ_{oo} = const. is thus permitted. One may suspect that only ratios of the γ_{ik} carry a physical meaning. Then this assumption becomes just a convention which is always possible. Leaving the unit of x^o undetermined so far we set:

$$\gamma_{oo} = \alpha. \tag{3}$$

Furthermore one shows that the following differential quantities are invariant under the transformation (2), namely[1]

$$d\theta = dx^o + \frac{\gamma_{oi}}{\gamma_{oo}}\, dx^i \ , \tag{4}$$

$$ds^2 = (\gamma_{ik} - \frac{\gamma_{oi}\gamma_{ok}}{\gamma_{oo}})dx^i dx^k \ . \tag{5}$$

In these expressions summation from 1 to 4 over indices appearing twice is understood. For those sums we shall, as usual, omit the summation symbol. The quantities $d\theta$ and ds are connected with the line element $d\sigma$ as

$$d\sigma^2 = \alpha d\theta^2 + ds^2 \ . \tag{6}$$

It follows from the invariance of $d\theta$ and γ_{oo} that the four γ_{oi} $(i \neq 0)$ for fixed x^o transform as the covariant components of a normal four vector. If x^o is also transformed, the gradient of a scalar has to be added. This means that the quantities

---

[1] cf. H.A. Kramers, Proc. Amsterdam 23, Nr. 7, 1922. In there is given an idea reminiscent of the following considerations together with a simple proof of the invariance of $d\theta$ and ds^2.

$$\frac{\partial\gamma_{oi}}{\partial x^k} - \frac{\partial\gamma_{ok}}{\partial x^i}$$

transform like the covariant components of the electromagnetic field tensor. From the point of view of invariant theory, the quantities γ_{oi} behave like the electromagnetic potential ϕ_i. Thus we assume

$$d\theta = dx^o + \beta\phi_i dx^i \quad , \tag{7}$$

i.e.,

$$\gamma_{oi} = \alpha\beta\phi_i \ , \qquad (i = 1,2,3,4) \tag{8}$$

where β is a constant and the ϕ_i are defined such that in rectangular Galileian coordinates

$$\left.\begin{aligned} (\phi_x, \phi_y, \phi_z) &= A \quad , \\ \phi_t &= -cV \ , \end{aligned}\right\} \tag{9}$$

where A stands for the usual vector potential, V is the usual scalar potential and c the velocity of light.

We should like to identify the differential form ds with the line element of usual relativity. We hence set

$$\gamma_{ik} = g_{ik} + \alpha\beta^2\phi_i\phi_k \tag{10}$$

where the g_{ik} are supposed to be chosen such that in rectangular Galileian coordinates

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2 \tag{11}$$

Thereby the quantities γ_{ik} have been reduced to known ones. The problem remains how to pose field equations for the γ_{ik} such that they yield for g_{ik} and ϕ_i the field equations of normal relativity in sufficient approximation. We do not want to go further into this difficult problem; we only intend to show that the usual field equations are readily combined from the point of view of five dimensional geometry. We form the invariant

$$P = \sum\gamma^{ik}\left[\frac{\partial\{{}^{i\mu}_{\ \mu}\}}{\partial x^k} - \frac{\partial\{{}^{ik}_{\ \mu}\}}{\partial x^\mu} + \{{}^{i\mu}_{\ \nu}\}\{{}^{k\nu}_{\ \mu}\} - \{{}^{ik}_{\ \mu}\}\{{}^{\mu\nu}_{\ \nu}\}\right] \tag{12}$$

where γ^{ik} denote the contravariant components of the five dimensional fundamental metric tensor and $\{{}^{rs}_{\ i}\}$ stands for Christoffel's three index symbols, i.e.,

$$\left\{ {rs \atop i} \right\} = \frac{1}{2} \Sigma \gamma^{i\mu} \left\{ \frac{\partial \gamma_{\mu r}}{\partial x^s} + \frac{\partial \gamma_{\mu s}}{\partial x^r} - \frac{\partial \gamma_{rs}}{\partial x^\mu} \right\} \qquad (13)$$

In the expression for P we take all quantities independent of x^o and $\gamma_{oo} = \alpha$.

Let us consider the integral

$$J = \int P\sqrt{-\gamma}\, dx^o dx^1 dx^2 dx^3 dx^4 \qquad (14)$$

taken over a closed region of five-dimensional space; γ stands for the determinant of the γ_{ik}.

We form δJ by varying the quantities γ_{ik} and $\frac{\partial \gamma_{ik}}{\partial x^\ell}$ while keeping the boundary values fixed. α is to be regarded as a constant. The variational principle

$$\delta J = 0 \qquad (15)$$

leads to the following equations

$$R^{ik} - \frac{1}{2} g^{ik} R + \frac{\alpha\beta^2}{2} S^{ik} = 0 \qquad (i,k = 1,2,3,4) \qquad (16a)$$

and

$$\frac{\partial \sqrt{-g}\, F^{i\mu}}{\partial x^\mu} = 0 \qquad (i = 1,2,3,4) \qquad (16b)$$

where we denote by R Einstein's curvature invariant, by R^{ik} the contravariant components of Einstein's curvature tensor, by g^{ik} the contravariant components of Einstein's fundamental tensor,by S^{ik} the contravariant components of the electromagnetic energy-momentum tensor, by g the determinant of g_{ik} and finally by $F^{i\mu}$ the contravariant components of the electromagnetic field tensor. If we put

$$\frac{\alpha\beta^2}{2} = \kappa \qquad (17)$$

where κ is the gravitational constant used by Einstein, we see that indeed equations (16a) are identical with the equations for the gravitational field of general relativity and equations (16b) with the generalized Maxwell equations of general relativity for a field point free of matter.[1]

---

[1] cf., e.g., W. Pauli, Relativitätstheorie, p. 719 and 724.

If we restrict ourselves to the schematic treatment of matter commonly used in the theory of electrons and relativity, we may obtain the usual equations for the case with matter in a similar manner. We replace P in (14) by

$$P + \kappa \sum \gamma_{ik} \Theta^{ik}$$

To define the Θ^{ik}, let us first consider the tensor

$$\theta^{ik} = \frac{dx^i}{d\ell} \frac{dx^k}{d\ell} \tag{18}$$

for an electron or a hydrogen nucleus, where dx^i denote the changes in position of the particle and $d\ell$ stands for a certain invariant differential. The Θ^{ik} are to be equal to the sum of the θ^{ik} for the various particles per unit volume. We again arrive at equations of the usual type which become identical with the usual field equations if we set

$$v_o \frac{d\tau}{d\ell} = \pm \frac{e}{\beta c}, \tag{19}$$

$$\frac{d\tau}{d\ell} = \begin{cases} \sqrt{M} \\ \sqrt{m} \end{cases}, \tag{20}$$

where generally

$$v_i = \sum \gamma_{i\mu} \frac{dx^\mu}{d\ell} \tag{21}$$

are the covariant components of the five dimensional velocity vector v^i and

$$v^i = \frac{dx^i}{d\ell} \tag{22}$$

Furthermore e stands for the electric elementary quantum and M respectively m for the masses of the hydrogen nucleus respectively the electron. Moreover

$$d\tau = \frac{1}{c} \sqrt{-ds^2}$$

is the differential of proper time.

The equations of motion for material particles and the equation of continuity follow, of course, in the usual manner from the field equations. The calculations yielding those results can be readily combined from our point of view. As can easily be seen, our field equations are equivalent to the following 14 equations:

$$P^{ik} - \frac{1}{2}\gamma^{ik} P + \kappa \Theta^{ik} = 0 \qquad (23)$$

($i,k = 0,1,2,3,4$, but not both zero), where the P^{ik} are the contravariant components of the contracted five dimensional curvature tensor (corresponding to the R^{ik}). The equations in question are now obtained by taking the divergence of (23). It follows that the electric particles move on five-dimensional geodesic lines which satisfy the conditions (12) and (20).[1] As can be seen immediately, those conditions are consistent with the equations of geodesic lines just because x^o does not appear in the γ_{ik}.

Here one should be reminded of the fact that there are really no sufficient reasons for the exact validity of Einstein's field equations. In any case it should be not without interest that all 14 field equations can be combined in such a simple manner from the point of view of Kaluza's theory.

§2. The Wave equation quantum theory. We now proceed to relate the theory of stationary states and the characteristic deviations from mechanics connected with it which appear in the new quantum theory to the five-dimensional theory of relativity. For this purpose let us consider the following differential equation which is to be related to five-dimensional space and can be viewed as a simple generalization of the wave equation:

$$\Sigma\, a^{ik} \left(\frac{\partial^2 U}{\partial x^i \partial x^k} - \Sigma \left\{ {ik \atop r} \right\} \frac{\partial U}{\partial x^r} \right) = 0 \,. \qquad (24)$$

The a^{ik} stand for the contravariant components of a five dimensional tensor which are supposed to be certain functions of the coordinates. Equation (24) exists independently of the system of coordinates.

Let us consider at first wave propagation determined by (24) which corresponds to the limiting case of geometrical optics. We arrive at this case, if we set

$$u = A\, e^{i\omega\Phi} \qquad (25)$$

---

[1] In this context the special values of $\frac{d\tau}{d\ell}$ are of no relevance. It only is important that $\frac{d\tau}{d\ell}$ = const.

and assume ω to be so large that only the terms proportional to ω^2 have to be taken into account. We then get

$$\Sigma\, a^{ik} \frac{\partial\Phi}{\partial x^i} \frac{\partial\Phi}{x^k} = 0 \,, \tag{26}$$

an equation which corresponds to the Hamilton-Jacobi partial differential equation of mechanics. Setting

$$p_i = \frac{\partial\Phi}{\partial x^i} \,, \tag{27}$$

the differential equations for the rays can, as is well known, be written in the following Hamiltonian form:

$$\frac{dp_i}{-\frac{\partial H}{\partial x^i}} = \frac{dx^i}{\frac{\partial H}{\partial p_i}} = d\lambda \,, \tag{28}$$

where

$$H = \frac{1}{2} \Sigma\, a^{ik} p_i p_k \,. \tag{29}$$

Furthermore it follows from (26)

$$H = 0 \,. \tag{30}$$

Another representation of those equations corresponding to the Lagrangean formulation follows from the fact that the rays can be regarded as geodesic null lines of the differential form:

$$\Sigma\, a_{ik} dx^i dx^k$$

where the a_{ik} are the quantities reciprocal to the a^{ik}, hence

$$\Sigma\, a_{i\mu} a^{k\mu} = \delta_i^k = \begin{cases} 1, & i=k \\ 0, & i=k \end{cases} \,. \tag{31}$$

If we take

$$\Sigma\, a_{ik} dx^i dx^k = \mu (d\theta)^2 + ds^2 \tag{32}$$

we can by an appropriate choice of the constant μ achieve that the equations for the rays become identical to the equations of motion of electric particles. To see this let us set

$$L = \frac{1}{2} \mu \left(\frac{d\theta}{d\lambda}\right)^2 + \frac{1}{2} \left(\frac{ds}{d\lambda}\right)^2 \,, \tag{33}$$

whence

$$p_0 = \frac{\partial L}{\partial \frac{dx^0}{d\lambda}} = \mu \frac{d\vartheta}{d\lambda} \tag{34}$$

and

$$p_i = \frac{\partial L}{\partial \frac{dx_i}{d\lambda}} = u_i \frac{d\tau}{d\lambda} + \beta p_0 \psi_i \qquad (i = 1,2,3,4) \;, \tag{35}$$

where $u_1 \ldots u_4$ denote the covariant components of the usual velocity vector.

Now the equations of rays are

$$\frac{dp_0}{d\lambda} = 0 \;, \tag{36a}$$

$$\frac{dp_i}{d\lambda} = \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x^i} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} + \beta p_0 \frac{\partial \psi_\nu}{\partial x^i} \frac{dx^\nu}{d\lambda} \qquad (i = 1,2,3,4) \;. \tag{36}$$

From

$$\mu d\vartheta^2 + ds^2 = \mu d\vartheta^2 - c^2 dt^2 = 0$$

it follows

$$\mu \frac{d\vartheta}{d\tau} = c\sqrt{\mu} \;. \tag{37}$$

As $\frac{d\vartheta}{d\lambda}$ and consequently $\frac{d\tau}{d\lambda}$ are constant due to (34) and (36a), we can choose λ such that

$$\frac{d\tau}{d\lambda} = \begin{cases} M & \text{for the hydrogen nucleus,} \\ m & \text{for the electron .} \end{cases} \tag{38}$$

Furthermore, in order to arrive at the usual equations of motion, we have to assume:

$$\beta p_0 = \begin{cases} +\frac{e}{c} & \text{for the hydrogen nucleus,} \\ -\frac{e}{c} & \text{for the electron.} \end{cases} \tag{39}$$

It then follows from (37)

$$\mu = \begin{cases} \frac{e^2}{\beta^2 M^2 c^4} & \text{for the hydrogen nucleus,} \\ \frac{e^2}{\beta^2 m^2 c^4} & \text{for the electron.} \end{cases} \tag{40}$$

Equations (35), (36) then completely agree with the usual equations of motion for electric particles in gravitational and electric fields. In particular, the quantities p_i defined according to (37) are identical with the generalized momenta defined in the usual way; this is important for the following considerations. As β is still arbitrary, let us set:

$$\beta = \frac{e}{c} . \tag{41}$$

It then simply follows that

$$p_o = \begin{cases} +1 & \text{for the hydrogen nucleus,} \\ -1 & \text{for the electron,} \end{cases} \tag{39a}$$

and

$$\mu = \begin{cases} \dfrac{1}{M^2c^2} & \text{for the hydrogen nucleus,} \\ \dfrac{1}{m^2c^2} & \text{for the electron.} \end{cases} \tag{40a}$$

As can be seen, we have to choose the positive sign for the square root in (37) in the case of the nucleus and the negative sign in the case of the electron. This is really not very satisfactory. On the other hand, the fact that one obtains, for a single value of μ, two different classes of rays which are, in a way, related to each other like the positive to the negative electric particles, could be understood as an indication that it may be possible to alter the wave equation such that the equations of motion for the two species of particles result from a single system of values for the coefficients. We do not intend to pursue this question further, rather we proceed to have a closer look at the wave equation following from (32) in the case of an electron.

Since we have taken $p_o = -1$ for the electron, according to (27) we have to set

$$\phi = -x_o + S(x^1,x^2,x^3,x^4) . \tag{42}$$

We arrive at de Broglies' theory if we look for standing waves which correspond to a certain value of ω consistent with the wave equation, while assuming that the waves propagate according to the laws of geometric optics. For this purpose we require the well-known theorem of conservation of phase which follows immediately from (28) and (30); namely

$$\frac{d\,\Phi}{d\,\lambda} = \Sigma \frac{\partial \Phi}{\partial \kappa^i} \frac{dx^1}{d\lambda} = \Sigma\, p_i \frac{\partial H}{\partial p_i} = 2H = 0 . \tag{43}$$

The phase is thus carried along by the wave. Let us consider now the simplest case where Φ can be separated into two terms, one of which depends only on one coordinate x, say, which oscillates back and forth periodically in time. Then a standing wave will be possible and it is characterized by the fact that a harmonic wave given at a certain instant by (25) coincides in phase after one period of x with that wave which results from the same solution (25) by inserting the new values x^o, x^2, x^3, x^4. Due to the conservation of phase the condition is simply

$$\omega \oint p dx = n \cdot 2\pi \tag{44}$$

where n is an integer. If we set

$$\omega = \frac{2\pi}{h} \tag{45}$$

where h stands for Planck's constant we arrive at the usual quantum condition for a separable coordinate. Analogous results hold, of course, for an arbitrary periodic system. The usual quantum theory of periodic systems thus is completely analogous to the treatment of interference phenomena under the assumption that the waves propagate according to the laws of geometrical optics. It may be stressed that due to (42) the relations (44), (45) are invariant under the coordinate transformations (2).

Let us consider equation (24) also in the case in which ω is not so large that we need to take into account only the terms quadratic in ω. We restrict ourselves to the simple case of an electrostatic field. Then we have in cartesian coordinates:

$$\begin{aligned} d\theta &= dx^o - eV\,dt\ , \\ ds^2 &= dx^2 + dy^2 + dz^2 - c^2 dt^2 . \end{aligned} \tag{46}$$

Thus we get:

$$H = \frac{1}{2}\left(p_x^2 + p_y^2 + p_z^2\right) - \frac{1}{2c^2}\left(p_t + eVp_o\right)^2 + \frac{m^2c^2}{2}\,p_o^2 \ . \tag{47}$$

In equation (24) we can neglect terms proportional to $\left\{ {ik \atop r} \right\}$, as the three-index symbols are in this case small quantities proportional to the gravitational constant κ. We obtain accordingly[1]

$$\Delta U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} - \frac{2eV}{c^2}\frac{\partial^2 U}{\partial t\,\partial x^o} + \left(m^2c^2 - \frac{e^2V^2}{c^2}\right)\frac{\partial^2 U}{\partial x^{o2}} = 0 \ . \tag{48}$$

As V depends only on x,y,z, we can make an Ansatz for U consistent with (42) and (45):

$$U = e^{-2\pi i(\frac{x^o}{h} - \nu t)} \psi(x,y,z) \;. \tag{49}$$

Inserted into (48) this yields

$$\Delta\psi + \frac{4\pi^2}{c^2h^2}(h\nu - eV)^2 - m^2c^4)\psi = 0 \;. \tag{50}$$

By letting

$$h\nu = mc^2 + E \tag{51}$$

we get Schrödinger's equation[2] whose standing waves, as is well known, correspond to values of E which are identical to the energy values calculated from Heisenberg's quantum theory. It can be seen that in the limiting case of geometrical optics E becomes equal to the mechanical energy defined in the usual manner. The condition on the frequency means according to (51), as Schrödinger has pointed out, that the light frequencies belonging to the system are equal to the differences formed with the various values of the frequency ν .

§3. Concluding remarks. The above considerations, like the papers by de Broglie, stem from the idea to utilize the analogy of mechanics and optics, apparent in Hamilton's method, for a deeper understanding of quantum phenomena. That there is a real physical sense to this analogy seems to be indicated by the similarity of the conditions for stationary states of an atomic system and interference phenomena in optics. As is well known, notions like point charge and material point are alien already to classical field theory. Furthermore, the hypothesis has been voiced frequently that material particles are to be considered as special solutions of the field equations determining the gravitational and electromagnetic field. It is suggestive to relate said analogy to this idea. According to this hypothesis it is not really strange that the motion of material particles resembles the propagation of waves. On the other hand the analogy in question is incomplete as long as one

---

[1] This equation differs from Schrödinger's equation, aside from the appearance of x^o, which is irrelevant for the applications anyway, in the way in which the time appears in (48). To support this form of a quantum equation it may be noted that it admits solutions in the case of V's depending harmonically on time which, as can easily be seen by perturbation methods, are related to Kramers' dispersion theory in a similar way as Schrödinger's solutions are related to the quantum theory of spectral lines. I am grateful to Dr. W. Heisenberg for this remark.

[2] Schrödinger, l.c.

considers wave propagation in a space of only four dimensions. This is already apparent in the variable velocity of material particles. The analogy can, however, be made complete, if one views the observed motion as a kind of projection onto space-time of a wave propagation taking place in a space of five dimensions. Mathematically speaking it means that the Hamilton-Jacobi equation can be regarded as an equation of characteristics not for a four dimensional but for a five dimensional wave equation. In this way one is led to Kaluza's theory.

Even though the introduction of a fifth dimension into our physical considerations may be a priori surprising, a radical modification of the geometry underlying the field equations is suggested in quite another way by quantum theory. It became, as is well known, less and less likely that quantum phenomena admit a unitary description in space-time, whereas the possibility to describe those phenomena by a system of five dimensional field equations cannot be excluded a priori.[1] Of course, the future has to decide whether behind all those indications of something possible there is something real. In any case it has to be stressed that the treatment attempted in this note as regards the field equations as well as the theory of stationary states has to be considered as quite preliminary. This is particularly apparent in the schematic treatment of matter surrounding equation (18)*, also in the fact that the two species of electric particles are described by different equations of Schrödinger's type. Also, the question remains entirely open, whether one can be content with the 14 potentials for a description of physical phenomena or whether Schrödinger's method entails the introduction of a new quantity of state.

I have worked on the ideas communicated in this note at the Institute of Physics, University of Michigan, Ann. Arbor, as well as at the Institute for Theoretical Physics here. At this point I should like to convey my warmest thanks to Prof. H.M. Randall and Prof. N. Bohr.

---

[1] Similar remarks by Prof. Bohr on several occasions have had a decisive influence on this note.

* The original text reads "on page 898" instead of "surrounding equation (18)".

THE ATOMICITY OF ELECTRICITY AS A QUANTUM THEORY LAW

By O. Klein

Reprinted from: Nature, Vol. 118, p.516, (1926)

IN the five-dimensional theory of the connexion between electromagnetism and gravitation first proposed by Th. Kaluza (*Sitzungsberichte d. Berl. Akad.* 1921, S. 766; see also O. Klein, *Zs. für Phys.*, 37, 8~~7~~5 895, 1926), the equations of motion of an electrified particle may be shown to be the equations of geodetics belonging to the following line element:

$$d\sigma = \sqrt{(dx^0 + \beta\phi_i dx^i)^2 + g_{ik}dx^i dx^k}, \qquad (1)$$

where x^1, x^2, x^3, x^4 are the co-ordinates of ordinary space time with the line element $g_{ik}dx^i dx^k$, while x^0 is a fifth co-ordinate, and the ϕ_i are the four co-variant components of the electromagnetic potential vector. If the constant β is given the value

$$\beta = \sqrt{2\kappa}, \qquad (2)$$

κ being the Einstein gravitational constant, the 14 field equations of the Einstein theory may, moreover, be simply expressed by means of the curvature tensor belonging to this line element.

Let now $d\tau$ be the differential of proper time belonging to a particle of mass m and charge e; then the Lagrange function L for the geodetics representing the motion of the particle may be given the form

$$L = \tfrac{1}{2}m\left(\frac{d\sigma}{d\tau}\right)^2. \qquad (3)$$

Defining momenta in the ordinary way by putting

$$p_i = \frac{\partial L}{\partial(dx^i/d\tau)} \qquad (i = 0, 1, 2, 3, 4), \qquad (4)$$

p_0 is seen to be constant along a geodetic, since x^0 does not appear in L. In addition to the equation expressing this constancy, the system belonging to (3) contains four equations which indeed become identical with the equations of motion of the particle if we put

$$p_0 = \frac{e}{\beta c}. \qquad (5)$$

With this choice of p_0 the momenta p_1, p_2, p_3, p_4 are further seen to agree with the ordinary definition of the momenta of an electrified particle.

Now the charge e, so far as our knowledge goes, is always a whole multiple of the electronic charge, so that we may write

$$p_0 = \frac{N\epsilon}{\beta c}, \qquad (6)$$

ϵ being the electronic charge and N a whole number, positive or negative. This formula suggests that the atomicity of electricity may be interpreted as a quantum theory law. In fact, if the five-dimensional space is assumed to be closed in the direction of x^0 with a period l, and if we apply the formalism of quantum mechanics to our geodetics, we shall expect p_0 to be governed by the following rule:

$$p_0 = N\frac{h}{l}, \qquad (7)$$

N being now a quantum number, which may be positive or negative according to the sense of motion in the direction of the fifth dimension, and h the constant of Planck. Comparing (7) with (6), and making use of the value (2) of β, we get for the period l:

$$l = \frac{hc\sqrt{2\kappa}}{\epsilon} = 0{\cdot}8 \times 10^{-30}\ \text{cm.} \qquad (8)$$

The small value of this length together with the periodicity in the fifth dimension may perhaps be taken as a support of the theory of Kaluza in the sense that they may explain the non-appearance of the fifth dimension in ordinary experiments as the result of averaging over the fifth dimension.

In a former paper (*Zs. für Phys., l.c.*) the writer has shown that the differential equation underlying the new quantum mechanics of Schrödinger can be derived from a wave equation of a five-dimensional space, in which h does not appear originally, but is introduced in connexion with a periodicity in x^0. Although incomplete, this result, together with the considerations given here, suggests that the origin of Planck's quantum may be sought just in this periodicity in the fifth dimension.

OSKAR KLEIN.

Copenhagen, September 3.

ON A GENERALIZATION OF KALUZA'S THEORY OF ELECTRICITY

By A. Einstein and P. Bergmann

(Received April 8, 1938)

Introduction

So far, two fairly simple and natural attempts to connect gravitation and electricity by a unitary field theory have been made, one by Weyl, the other by Kaluza. Furthermore, there have been some attempts to represent Kaluza's theory formally so as to avoid the introduction of the fifth dimension of the physical continuum. The theory presented here differs from Kaluza's in one essential point; we ascribe physical reality to the fifth dimension whereas in Kaluza's theory this fifth dimension was introduced only in order to obtain new components of the metric tensor representing the electromagnetic field. Kaluza assumes the dependence of the field variables on the four coördinates x^1, x^2, x^3, x^4 and not on the fifth coördinate x^0 when a suitable coördinate system is chosen.

It is clear that this is due to the fact that the physical continuum is, according to our experience a four dimensional one. We shall show, however, that it is possible to assign some meaning to the fifth coördinate without contradicting the four dimensional character of the physical continuum.

In the first chapter of our paper we present Kaluza's original theory; in the second, its new generalization. This is done in order to make the reading easier.

In the appendix we simplify the derivation of the field equations by generalizing the tensor calculus for the case of tensor densities.

I. The Kaluza Theory[1]

1. We consider a five dimensional space (4 + 1 dimensions) with the metric

$$d\sigma^2 = \gamma_{\mu\nu} dx^\mu dx^\nu. \tag{1}$$

2. We assume that the space is "cylindrical" with respect to a definite vector field. Analytically this means: There exists a contravariant vector A^ν so that if τ is an infinitesimally small constant, then

$$\delta x^\nu = \tau \cdot A^\nu \tag{2}$$

is an infinitesimal displacement of the space points. The element must not change its length given by (1) if its endpoints are displaced according to (2). This is expressed by the equation:

---

[1] Contained in part in O. Klein, Zeitschrift f. Phys., **37** (1926), 895–906, A. Einstein, Sitzungsberichte d. Preuss. Akademie d. Wissenschaften, Physik.-mathemat. Klasse, 1927, 23–30.

$$\gamma_{\mu\nu,\sigma}A^{\sigma} + \gamma_{\mu\sigma}A^{\sigma}_{,\nu} + \gamma_{\nu\sigma}A^{\sigma}_{,\mu} = 0. \tag{3a}$$

The invariant character of this equation can easily be shown. Putting

$$\gamma_{\mu\sigma}A^{\sigma} = A_{\mu}, \tag{3b}$$

we get

$$A_{\mu,\nu} + A_{\nu,\mu} - A^{\sigma}\{\gamma_{\mu\sigma,\nu} + \gamma_{\nu\sigma,\mu} - \gamma_{\mu\nu,\sigma}\} = 0, \tag{3c}$$

or

$$\{A_{\mu,\nu} - A_{\sigma}\Gamma^{\sigma}_{\mu\nu}\} + \{A_{\nu,\mu} - A_{\sigma}\Gamma^{\sigma}_{\nu\mu}\} = 0, \tag{3d}$$

where $\Gamma^{\lambda}_{\iota\kappa}$ are the Christoffel symbols of the second kind. Or, in the language of the absolute tensor calculus we have:

$$A_{\mu;\nu} + A_{\nu;\mu} = 0. \tag{3}$$ [2]

It follows from (3) that $\gamma_{\mu\nu}A^{\mu}A^{\nu}$ (that is the norm of A^{μ}) is constant along the lines to which A^{μ} is a tangent. By multiplying (3) by $A^{\mu}A^{\nu}$ we obtain:

$$A^{\nu}A^{\mu}A_{\mu;\nu} + A^{\mu}A^{\nu}A_{\nu;\mu} = A^{\nu}(A^{\mu}A_{\mu})_{,\nu} = 0. \tag{4}$$

3. Kaluza's theory imposes another property on the vector field A^{μ}, besides the one expressed in (3): The lines to which the A^{μ} are tangents—the "A-lines"—have to be geodesics. Analytically this means: The lines defined by

$$\frac{dx^{\nu}}{d\sigma} = \lambda A^{\nu} \tag{5}$$

(where $1/\lambda^2$ is equal to the norm of A) satisfy the equation

$$\frac{d^2x^{\nu}}{d\sigma^2} + \Gamma^{\nu}_{\alpha\beta}\frac{dx^{\alpha}}{d\sigma}\frac{dx^{\beta}}{d\sigma} = 0, \tag{6a}$$

or, according to (5):

$$\frac{\partial \ln \lambda}{\partial x^{\alpha}}A^{\alpha}A^{\nu} + A^{\alpha}A^{\nu}_{;\alpha} = 0. \tag{6b}$$

The first term vanishes because of (4) $\left(\frac{\partial \ln \lambda}{\partial x^{\alpha}}\cdot A^{\alpha}\right.$ is the derivative of a function of the norm of A in the direction of $\left.A\right)$. Therefore:

$$A^{\alpha}A^{\nu}_{;\alpha} = 0, \tag{6c}$$

---

[2] Killing's equation.

or

(6d)
$$A^{\alpha}A_{\nu;\alpha} = 0.$$

From (3) follows:

(6)
$$A^{\alpha}A_{\alpha;\nu} = \tfrac{1}{2}(A^{\alpha}A_{\alpha})_{,\nu} = 0.$$

This means that the norm of A is constant not only along the A-lines but in the whole space. We characterize, in this way, the space structure in Kaluza's theory.

Remark: From the "cylindrical" character of the space there follows the existence of a vector A the *symmetrical* derivatives of which vanish. But there remain the *antisymmetrical* derivatives of A_{μ} which represent, in Kaluza's theory, the electromagnetic field.

The Special Coördinate System

We consider a four dimensional surface cutting each of the A-lines once and only once. We introduce on this surface 4 coördinates x^{a} ($a = 1 \cdots 4$) and assume x^{0} equal zero on this surface. Through each point of the surface passes an A-line on which one of the two directions is chosen as positive; this choice is to be the same on all A-lines. As x^{0}-coördinate we choose the distance along an A-line calculated with the help of (1), starting from the original four dimensional surface with $x^{0} = 0$, the chosen direction determining the sign of x^{0}.

Therefore, on the x^{0}-line:

(7a)
$$x^{0} = \int_{0}^{x^{0}} \sqrt{\gamma_{00}}\, dx^{0},$$

and therefore in the whole space,

(7)
$$\gamma_{00} = 1.$$

As the absolute value of A is always 1, we have

(8a)
$$\gamma_{00}A^{0}A^{0} = 1,$$

or,

(8)
$$A^{0} = 1,$$

since, because of the choice of our coördinate system, $A^{1} = A^{2} = A^{3} = A^{4} = 0$.

From this and (3a), expressing the condition of cylindricity, we have

(9)
$$\gamma_{\mu\nu,0} = 0.$$

As generally

$$A_{\mu} = \gamma_{\mu\nu}A^{\nu},$$

we have in the special coördinate system,

$$A_m = \gamma_{0m}, \qquad A_0 = \gamma_{00} = 1. \tag{10}$$

None of the components of A_μ depend on x^0 because of (9). Therefore in the antisymmetrical derivatives

$$A_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu}$$

all components bearing one index zero vanish identically.

In our special coördinate system the space structure is described by 10 functions γ_{mn} and 4 functions $\gamma_{0m}(= A_m)$, which do not depend on x^0. (Here and later the latin indices run only from 1 to 4.) This is, therefore, a four dimensional description of the space. But the variables introduced are not, from the point of view of our description, the most natural ones, as will be shown soon.

We ask: What coördinate transformations are still allowed preserving our special coördinate system?

a. The coördinates x^a have been chosen arbitrarily on the four dimensional surface (with $x^0 = 0$). Therefore the following transformation is still allowed:

$$\begin{aligned} \bar{x}^a &= \bar{x}^a(x^1 \cdots x^4), \\ \bar{x}^0 &= x^0. \end{aligned} \tag{11}$$

We call it the *Four-transformation.*

b. The four dimensional surface (with $x^0 = 0$) was chosen arbitrarily. It can be replaced by another surface without changing the x^a-coördinates corresponding to the A-lines. The appropriate transformation is:

$$\begin{aligned} \bar{x}^a &= x^a, \\ \bar{x}^0 &= x^0 + f(x^1 \cdots x^4). \end{aligned} \tag{12}$$

We call it the *Cut-transformation.*

The γ_{mn} behave, with respect to a four-transformation, as a four dimensional tensor and $A_n(= \gamma_{0n})$ as a four-vector. With respect to the cut-transformation we have

$$\bar{A}_m = A_m - \frac{\partial f}{\partial x^m}, \tag{13}$$

$$\bar{\gamma}_{mn} = \gamma_{mn} - \gamma_{m0}\frac{\partial f}{\partial x^n} - \gamma_{n0}\frac{\partial f}{\partial x^m} + \frac{\partial f}{\partial x^m}\frac{\partial f}{\partial x^n}. \tag{14}$$

Instead of the γ_{mn} which are not invariant with respect to cut-transformations (according to (14)) we introduce new functions g_{mn} by the following consideration: Let two infinitesimally close A-lines L and L' pass through two infinitesimally close points P and P'. There exists between L and L' an extremal distance because of the existence of a metric. This distance is determined by

$$g_{\mu\nu} = \gamma_{\mu\nu} - A_\mu A_\nu. \tag{15}$$

The $g_{\mu\nu}$ defined in this way are components of a five dimensional tensor. In the special coördinate system only those of its components which have no zero index do not vanish, i.e. the g_{mn} are its only components. The g_{mn} behave like a four dimensional tensor with respect to the four-transformation and are invariant with respect to cut-transformations.

It is, therefore, quite natural to choose g_{mn} and A_m as field variables in the special coördinate system; the g_{mn} are the components of the physical metrical tensor. The antisymmetrical derivatives of A_m, that is $A_{m,n} - A_{n,m}$, are invariant with respect to the cut-transformation, but not the components A_m themselves. This corresponds to the fact that the electromagnetic potentials are defined only up to additive terms which are gradients of an arbitrary function.

This result can be summarized: The five dimensional space structure is equivalent to a four dimensional one with a metric (g_{mn}) and a vector field (A_m), determined only up to an additive arbitrary gradient.

Kaluza's roundabout way of introducing the five dimensional continuum allows us to regard the gravitational and electromagnetic fields as a unitary space structure. This result is essential. At first sight it could be argued: Is it really a step forward to introduce a five dimensional metric and a vector for which we assume some arbitrary restrictions instead of a four dimensional metric and a four-vector? But to be just to Kaluza's theory it should be born in mind that the one really arbitrary step is taken only when the five dimensional representation of the four dimensional space is assumed. If, however, this is done, then the introduction of a five-vector is merely a consequence, necessary because of the four dimensional character of the empirical space. The representation of the electromagnetic field by a potential vector, achieved in this way, is certainly not a trivial result.

But apart from this point, the result is rather disappointing. Kaluza's aim was undoubtedly to obtain some new physical aspect for gravitation and electricity by introducing a unitary field structure. This end was, however, not achieved.

The field equations of the theory can be derived for instance with the help of a variational principle. This means a choice of an action function $\mathfrak{H}$, which must be a scalar density, i.e. $\frac{1}{\sqrt{g}} \cdot \mathfrak{H}$ must be an invariant with respect to the four- and cut-transformations. Therefore, in our special coördinate system, the invariant has to be formed of the g_{mn} and A_m as well as of such derivatives of these quantities which do not change if A_m is replaced by $A_m - \partial f/\partial x^m$. We can conclude, therefore, that A_m can enter the action function only through their antisymmetric derivatives. If the Hamiltonian $\mathfrak{H}$ has to contain only second derivatives of the potentials g_{mn}, A_m, or expressions of the second order containing only first derivatives, then the most general action function is of the form

$$\mathfrak{H} = \sqrt{g}(R + c \cdot M), \tag{16}$$

where R is the curvature invariant and M the Maxwell invariant.

Many fruitless efforts to find a field representation of matter free from singularities based on this theory have convinced us, however that such a solution does not exist.[3]

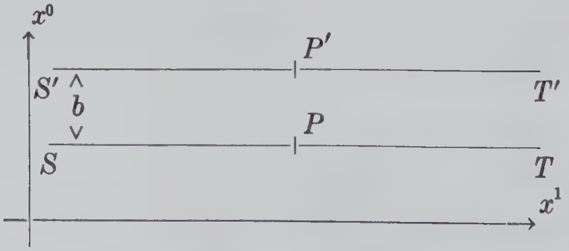
II. Generalization of Kaluza's Theory

General Remarks about the Space Structure

If Kaluza's attempt is a real step forward, then it is because of the introduction of the five dimensional space. There have been many attempts to retain the essential formal results obtained by Kaluza without sacrificing the four dimensional character of the physical space. This shows distinctly how vividly our physical intuition resists the introduction of the fifth dimension. But by considering and comparing all these attempts one must come to the conclusion that all these endeavors did not improve the situation. It seems impossible to formulate Kaluza's idea in a simple way without introducing the fifth dimension.

We have, therefore, to take the fifth dimension seriously although we are not encouraged to do so by plain experience. If, therefore, the space structure seems to force acceptance of the five dimensional space theory upon us we must ask whether it is sensible to assume the rigorous reducibility to four dimensional space. We believe that the answer should be "no," provided that it is possible to understand, in another way, the quasi-four dimensional character of the physical space by taking as a basis the five dimensional continuum and to simplify hereby the basic geometrical assumptions.

To make the underlying ideas clearer we consider a two dimensional space (instead of the five dimensional one), that is approximately a one dimensional continuum (instead of a four dimensional one).



---

[3] We tried to find a rigorous solution of the gravitational equations, free from singularities, by taking into account the electromagnetic field. We thought that a solution of a rotational-symmetrical character could, perhaps, represent an elementary particle. Our investigation was based on the theory of "bridges" (Einstein and Rosen, Phys. Rev., **48:** 73 (1935). We convinced ourselves, however, that no solution of this character exists.

The space considered is the narrow strip limited by the lines ST and $S'T'$, The coördinates are x^0 and x^1, and, for the sake of simplicity, the Euclidean character of the space is assumed. We imagine the strip curved into a tube so that ST coincides with $S'T'$. Every point P on ST coincides in this way with a certain point P' on $S'T'$, so that a small cylindrical surface is formed. If the width of the strip, that is, the circumference of the cylinder (denoted by b), is small, and if a continuous and slowly changing function $\varphi(x^0, x^1)$ is given, that is if $b \cdot \partial\varphi/\partial x^a$ is small compared with φ, then the values of φ belonging to the points on the segment PP' differ from each other very slightly and φ can be regarded, approximately, as a function of x^1 only. This reduction of the number of dimensions of the space was achieved because the space is closed in the x^0-dimension and the "width" is very small. But this quasi-one dimensional character of the space does not exist if the function φ varies too rapidly.

The assumption concerning the Euclidean character is certainly unessential for our consideration. It may be essential for our argument that some metric is assumed so that it is possible to speak of a circumference in the x^0-direction, but in any case we are interested only in spaces with a Riemannian metric.

It is more convenient to consider, in the following, instead of a space "closed" in the x^0-direction[4] a space "periodic" in the x^0-direction. The "periodic" and the "closed" character are equivalent if the corresponding points P, P', P'', $\cdots$ are regarded as "the same" point.

By putting in our picture the four dimensional continuum $x^1 \cdots x^4$ instead of the one dimensional continuum x^1 we obtain a model of the physical space which will form the basis of our future considerations.

The most essential point of our theory is the replacing of the hypothesis **2**, of rigorous cylindricity by the assumption that space is closed (or periodic) in the x^0-direction.

The Space Structure

We shall characterize the space according to our previous remarks concerning the modified form of Kaluza's theory.

1. We consider a five dimensional space with the metric (1).

2. The space is closed with respect to one dimension. This closed space will be represented by a space which is open and periodic with respect to this dimension. A point P of the physical space will, therefore, be represented by an infinite number of points P, P', P'', $\cdots$.

REMARK: This postulate replaces the cylindricity condition in Kaluza's original theory.

3. Through every space point there passes a geodesic line closed in itself and free from singularities. Or, formulated for the "periodic" space: If two points P, P' corresponding to each other, but otherwise arbitrary, are connected by

---

[4] This expression is not quite clear but it may be used here in a chapter aiming only at a general orientation.

the geodesic line,[5] then this line also passes through all other points corresponding to P, i.e. through P'', P''' etc.

This axiom is analogous to the axiom 3. of the original Kaluza theory. The geodesic lines introduced here correspond to the A-lines in Kaluza's theory and we shall again call them the A-lines.

We again introduce the special coördinate system. We consider a four dimensional surface $x^0 = 0$ cutting all A-lines once and only once. The four coördinates $x^1 \cdots x^4$ on this surface determine a point on this surface as well as an A-line through this point. Any point P on this A-line is determined by the distance (measured along this line with the help of (1)) between $\overset{\circ}{P}$ (that is the intersecting point of the A-line and the surface $x^0 = 0$) and P. For points on the "positive" side of the surface $x^0 = 0$ we have

$$x^0 = \frac{1}{b} \int_{\overset{\circ}{P}}^{P} d\sigma, \tag{17}$$

where b is

$$b = \int^{P'} d\sigma. \tag{17a}$$

Here P, P' are two corresponding, consecutive points lying on the A-line. Therefore, b depends on $x^1 \cdots x^4$ only. By this convention the difference of the x^0-coördinates of two corresponding, consecutive points (which, of course, lie on the same x^0-line) equals 1.

By the assumptions thus made five coördinates x^0, $x^1 \cdots x^4$ correspond to every space point.

Lines with constant $x^1 \cdots x^4$ are geodesic lines. On these lines we have:

$$\frac{dx^1}{d\sigma} = \frac{dx^2}{d\sigma} = \frac{dx^3}{d\sigma} = \frac{dx^4}{d\sigma} = 0, \qquad \frac{dx^0}{d\sigma} = \frac{1}{b}, \qquad \frac{d^2x^0}{d\sigma^2} = 0. \tag{18a}$$

Therefore, the equation of the geodesic line gives

$$\Gamma^{\alpha}_{00} = 0. \tag{18b}$$

From this follows:

$$\begin{bmatrix} \alpha \\ 00 \end{bmatrix} = 0, \tag{18c}$$

or

$$\gamma_{00,0} = 0 \tag{18d}$$

and

$$2\gamma_{\alpha 0,0} - \gamma_{00,\alpha} = 0. \tag{18}$$

---

[5] We do not consider the existence of more than one geodesic connection.

Equation (18d) does not contain anything new because we have on the A-lines:

$$d\sigma^2 = b^2 dx^{0 2} \tag{19a}$$

and

$$\gamma_{00} = b^2 \tag{19}$$

which is independent of x^0.

We integrate (18) over one period along the A-line. The integral of the first term on the left hand side in (18) gives zero because of the periodic character of the space. Because of (17) and (19) we obtain:

$$\frac{\partial}{\partial x^a} \int_{P'}^{P''} \gamma_{00}\, dx^0 = \frac{\partial}{\partial x^a} \left\{ b \int_{P'}^{P''} d\sigma \right\} = \frac{\partial}{\partial x^a} (b^2) = 0. \tag{20}$$

Therefore b and γ_{00} are independent of $x^1 \cdots x^4$ too, i.e. constant in the whole space. We may write:

$$\begin{cases} \gamma_{00} = 1, \\ b = 1. \end{cases} \tag{21}$$

In a strict sense we were not justified in choosing the value for γ_{00} as we have already chosen the coördinate difference Δx^0 of corresponding, consecutive points to be equal 1. We take this fact into account by putting Δx^0 not 1, but some small quantity λ. Then, because of (21), λ is simultaneously the metric distance $P'P''$.

Furthermore, from (18) follows because of (21)

$$\frac{\partial}{\partial x^0} \gamma_{a0} = 0. \tag{22}$$

Therefore, in the generalized theory also, the γ_{0a} (but not the γ_{ab}) are independent of x^0. We can introduce, in the generalized theory as well as before, the field of contravariant unit vectors A^μ the components of which are, in the special coördinate system, 0, 0, 0, 0, 1. Therefore again

$$A_a = A^\beta \gamma_{\beta a} = \gamma_{0a}. \tag{22a}$$

We can also introduce as the metric tensor instead of the

$$g_{\alpha\beta} = \gamma_{\alpha\beta} - A_\alpha A_\beta, \tag{23}$$

of which only those components which have no index zero do not vanish. There exists, in this sense, here also a four dimensional metrical tensor g_{ab}. *Its components are, however, in general periodic functions of* x^0. The whole difference between the new theory and Kaluza's lies just in this fact. (A_m in both theories do not depend on x^0.)

Having characterized the space structure we shall turn later, after some mathematical preparations, to the formulation of field equations which can be derived from a variational principle in the space considered. We shall again demand that the Hamiltonian be composed additively of terms containing either second derivatives (linearly) or products of two first derivatives.

Tensor Analysis with Respect to the Special Coördinate System

From our definition of the special coördinate system it follows that again there exists the freedom of "four-transformations" and "cut-transformations." In order to avoid the formal difficulties of general coördinate systems we have to investigate the transformation properties of the quantities formed with respect to the four- and cut-transformations. Definition of the contravariant four-vector: Four quantities a^s (being functions of x^0, $x^1 \cdots x^4$) form a contravariant vector if, with respect to four-transformations, the following transformation law is valid:

$$\bar{a}^s = \frac{\partial \bar{x}^s}{\partial x^t} a^t, \tag{24}$$

and if the a^s are invariant with respect to cut-transformations. The definition of a covariant vector is quite analogous. The definition of a tensor also follows from this.

Therefore g_{mn} is a tensor. However $A_m (= \gamma_{0m})$ is not a vector as the components transform, with respect to a cut-transformation, according to (13). The $A_{mn} = A_{m,n} - A_{n,m}$ are, however components of a tensor.

The tensor algebra is the same as the normal four dimensional one and does not need to be carried out here.

We now direct our attention to the formation of new tensors by differentiation.

1. For both four- and cut-transformations we have

$$\frac{\partial}{\partial \bar{x}^0} = \frac{\partial}{\partial x^0}. \tag{25}$$

Therefore: From every tensor a new one may be formed by differentiation with respect to x^0.

2. Covariant differentiation with respect to x^a: If ψ is a scalar then we have for cut-transformations:

$$\frac{\partial \psi}{\partial \bar{x}^a} = \frac{\partial \psi}{\partial x^a} - \frac{\partial \psi}{\partial x^0} f_{,a}. \tag{26a}$$

From this equation and (13), by eliminating $f_{,a}$, we get that

$$\frac{\partial \psi}{\partial x^a} - A_a \frac{\partial \psi}{\partial x^0}$$

is invariant with respect to the cut-transformations. Therefore the operator

$$\left(\frac{\partial}{\partial x^a} - A_a \frac{\partial}{\partial x^0}\right), \tag{26}$$

acting on a tensor, leaves its components invariant with respect to cut-transformations.

On the other hand, we know from the absolute differential calculus that the expression

(27a) $$B_{s,a} - \tfrac{1}{2}B_l g^{lm}(g_{ms,a} + g_{am,s} - g_{as,m})$$

(B_s being an arbitrary covariant vector) has a tensor character with respect to four- (but not cut-) transformations. By replacing $\dfrac{\partial}{\partial x^a}$ in this expression everywhere by the operator (26) we add new members which again have a tensor character with respect to four-transformations, this being true for $B_{s,0}$, $g_{mn,0}$, A_s. The expression obtained in this way certainly has a tensor character with respect to four-transformations. Since it is also invariant with respect to cut-transformations it satisfies our definition of a tensor.

We can, therefore, define an invariant differentiation in the following way:

(27) $$B_{s;a} = B_{s,a} - A_a B_{s,0} - B_l \Gamma^l_{sa},$$

where

(27b) $$\Gamma^l_{sa} = \tfrac{1}{2}g^{lm}[(g_{ms,a} - A_a g_{ms,0}) + (g_{ma,s} - A_s g_{ma,0}) - (g_{as,m} - A_m g_{as,0})].$$

In an analogous way we have for a contravariant vector with the same Γ^l_{sa}:

(28) $$B^s_{;a} = B^s_{,a} - A_a B^s_{,0} + B^l \Gamma^s_{la}.$$

From this a corresponding rule for more general tensors follows immediately. The absolute differentiation defined thus satisfies the rule of differentiation of products.

By straightforward calculation one obtains

(29) $$g_{mn;a} = g^{mn}{}_{;a} = 0.$$

3. The curvature tensor: By a straightforward calculation we find:

(30) $$B_{s;a;b} - B_{s;b;a} = -B_l R^l_{.sab} - B_{s,0} A_{ab},$$ [6]

where

(30a) $$R^l_{.sab} = (\Gamma^l_{sa,b} - A_b \Gamma^l_{sa,0}) - (\Gamma^l_{sb,a} - A_a \Gamma^l_{sb,0}) - \Gamma^l_{am}\Gamma^m_{sb} + \Gamma^l_{bm}\Gamma^m_{sa}.$$

From (30) follows, because of the tensor character of the last term, that $R^l_{.sab}$ is a tensor (the curvature tensor) which is antisymmetric with respect to its last two indices.

4. The commutation rule for the operators "$_{,0}$" and "$_{;a}$" is

(31) $$(u_{m,0})_{;a} - (u_{m;a})_{,0} = u_s \Gamma^s_{am,0}.$$

The derivatives of the Γ^s_{am} with respect to x^0 have, therefore, tensor character

---

[6] $A_{ab} = A_{a,b} - A_{b,a}$.

Variational Principle and Field Equations

Starting from the formalism developed here we shall formulate the most general field equations satisfying the following conditions:

1. The field equations should be derived from a variational principle;
2. The action function should consist exclusively of terms containing either second derivatives linearly or products of two derivatives of the first order.

These are the tensors from which such an action function could be built algebraically:

$$R^{i}_{.klm}\,,\; A_{mn}\,,\; g_{mn}\,,\; g_{mn,0}\,,\; g_{mn,00}\,,\; A_{mn,0}\,,\; A_{mn;s}.$$

Some of the invariants formed from these tensors can, by partial integration of the Hamiltonian integral be reduced to the others. There remain only the following invariants which are, in this sense, independent of each other:

$$(32)\quad \begin{cases} \underset{1}{H} = R^{i}_{.klm}\delta^{l}_{i}g^{km} = R_{km}g^{km} = R, \qquad \underset{2}{H} = A_{mn}A_{st}g^{ms}g^{nt} = A_{mn}A^{mn}, \\ \underset{3}{H} = g^{mn}{}_{,0}\,g_{mn,0}\,, \qquad \underset{4}{H} = g^{mn}g_{mn,0}\,g^{rs}g_{rs,0}\,. \end{cases}$$

The most general variational principle, satisfying our conditions formulated above, is therefore:

$$(33)\quad \delta\int \mathfrak{H}\,dx^{0}dx^{1}dx^{2}dx^{3}dx^{4} = 0, \qquad \mathfrak{H} = \sqrt{-g}\Big(\underset{1}{\alpha}\underset{1}{H} + \underset{2}{\alpha}\underset{2}{H} + \underset{3}{\alpha}\underset{3}{H} + \underset{4}{\alpha}\underset{4}{H}\Big),$$

where $\underset{1}{\alpha} \cdots \underset{4}{\alpha}$ are arbitrary constants. The variation has to be carried out with respect to the g^{mn} and the A_s. The integral should be taken over an arbitrary world region of the coördinates $x^1 \cdots x^4$ and over one period of the coördinates x^0. The latter restriction is necessary because the δA_s must be chosen independent of x^0. (Compare (22)). Therefore the contributions from the boundary of the region do not enter into the usual operations performed on the action function only if the integral is taken over exactly one period of x^0. The calculations, which may be performed with the help of the methods quoted in the appendix, give:

$$(34)\quad \begin{cases} \delta\int \underset{1}{\mathfrak{H}}\,dx^{0}\cdots dx^{4} = \int \{[\tfrac{1}{2}R_{kl} + \tfrac{1}{2}R_{lk} - \tfrac{1}{2}g_{kl}R]\,\delta g^{kl} \\ \qquad\qquad + [g^{mn}\Gamma^{s}_{mn,0} - g^{ms}\Gamma^{n}_{mn,0}]\,\delta A_s\}\sqrt{-g}\,dx^{0}\cdots dx^{4}, \\ \delta\int \underset{2}{\mathfrak{H}}\,dx^{0}\cdots dx^{4} = \int \{[2A_{km}A_{l.}^{\;m} - \tfrac{1}{2}g_{kl}A_{mn}A^{mn}]\,\delta g^{kl} \\ \qquad\qquad - [4A^{st}{}_{;t}]\,\delta A_s\}\sqrt{-g}\,dx^{0}\cdots dx^{4} \\ \delta\int \underset{3}{\mathfrak{H}}\,dx^{0}\cdots dx^{4} = \int [-2g_{kl,00} + 2g^{rs}g_{kr,0}\,g_{ls,0} - g^{rs}g_{rs,0}\,g_{kl,0} \\ \qquad\qquad - \tfrac{1}{2}g^{rs}{}_{,0}\,g_{rs,0}\,g_{kl}]\,\delta g^{kl}\sqrt{-g}\,dx^{0}\cdots dx^{4}, \\ \delta\int \underset{4}{\mathfrak{H}}\,dx^{0}\cdots dx^{4} = \int g_{kl}[\tfrac{1}{2}(g^{mn}g_{mn,0})^{2} + 2g^{mn}g_{mn,00} \\ \qquad\qquad + 2g^{mn}{}_{,0}\,g_{mn,0}]\,\delta g^{kl}\sqrt{-g}\,dx^{0}\cdots dx^{4}. \end{cases}$$

Here the δg^{kl} are arbitrary functions of the coördinates $x^0 \cdots x^4$. The sum of the coefficients of δg^{kl}, multiplied by their $\underset{i}{\alpha}$, must, therefore, vanish. However, the δA_s are, as we mentioned before, functions of $x^1 \cdots x^4$ only, but independent of x^0. Therefore, in this case, only the integral of the sum of the coefficients of δA_s, (multiplied by their $\underset{i}{\alpha}$), taken over one period of x^0, vanishes. The fourteen field equations derived in this way from the variational principle (33), take the form:

$$(35)\quad \begin{cases} \underset{1}{\alpha}(\tfrac{1}{2}R_{kl} + \tfrac{1}{2}R_{lk} - \tfrac{1}{2}g_{kl}R) + \underset{2}{\alpha}(2A_{km}A_{l.}^{\ m} - \tfrac{1}{2}g_{kl}A_{mn}A^{mn}) + \\ + \underset{3}{\alpha}(-2g_{kl,00} + 2g^{rs}g_{kr,0}g_{ls,0} - g^{rs}g_{rs,0}g_{kl,0} - \tfrac{1}{2}g^{rs}{}_{,0}g_{rs,0}g_{kl}) + \\ + \underset{4}{\alpha}g_{kl}[\tfrac{1}{2}(g^{mn}g_{mn,0})^2 + 2g^{mn}g_{mn,00} + 2g^{mn}{}_{,0}g_{mn,0}] = 0 = \dfrac{1}{\sqrt{-g}}\mathfrak{G}_{kl}, \end{cases}$$

$$(36)\quad \int \{\underset{1}{\alpha}(g^{mn}\Gamma^s_{mn,0} - g^{ms}\Gamma^n_{mn,0}) - 4\underset{2}{\alpha}A^{st}_{;t}\}\sqrt{-g}\,dx^0 = 0 = \int \mathfrak{J}^s\,dx^0.$$

Finally we shall derive the identities which the field equations satisfy. As it is customary, we shall find these identities by expressing analytically the invariance of the Hamiltonian integral with respect to an infinitesimal coördinate transformation. If we introduce for δg^{mn} and δA_s the variations produced by an infinitesimal coördinate transformation, then the variation of the integral vanishes. It should be noted that the variation is not to be carried out with the space points fixed, but with the coördinate values fixed. Denoting the infinitesimal coördinate transformations by

$$(37a)\quad \begin{cases} x^a = \bar{x}^a + \xi^a(\bar{x}^1 \cdots \bar{x}^4) \\ x^0 = \bar{x}^0 + \xi^0(\bar{x}^1 \cdots \bar{x}^4) \end{cases}$$

we obtain for δg^{mn} and δA_s :

$$(37b)\quad \begin{cases} \delta g^{kl} = -g^{sl}\xi^k_{,s} - g^{sk}\xi^l_{,s} + g^{kl}{}_{,s}\xi^s + g^{kl}{}_{,0}\xi^0, \\ \delta A_s = A_r\xi^r_{,s} + \xi^0_{,s} + A_{s,r}\xi^r. \end{cases}$$

These expressions can easily be brought into a form revealing more clearly their tensor character:

$$(37c)\quad \begin{cases} \delta g^{kl} = -g^{sl}\xi^k{}_{;s} - g^{sk}\xi^l_{;s} + g^{kl}{}_{,0}(A_r\xi^r + \xi^0), \\ \delta A_s = (A_r\xi^r + \xi^0)_{;s} + A_{sr}\xi^r. \end{cases}$$

The expression $(A_r\xi^r + \xi^0)$ is an inner product of two five dimensional vectors, therefore a scalar. After performing some partial integrations (using the methods explained in the appendix) we obtain:

$$(37d)\quad \delta\int \mathfrak{H}dx^\alpha = \int \{[\mathfrak{G}_{k.\,;s}^{\ s} + A_{sk}\mathfrak{J}^s]\xi^k + [\mathfrak{G}_{rs}g^{rs}{}_{,0} - \mathfrak{J}^s_{;s}](A_r\xi^r + \xi^0)\}\,dx^\alpha = 0.$$

As ξ^k, ξ^0 are arbitrary functions of $x^1 \cdots x^4$, but independent of x^0 the integral (37d) will vanish identically if and only if the integrals of the square bracket expressions over one period of x^0 equal zero:

$$(37) \qquad \begin{cases} 2\int \mathfrak{G}_{k.\,;s}^{\ s}\, dx^0 + A_{sk} \int \mathfrak{J}^s\, dx^0 = 0, \\ \int \mathfrak{G}_{rs}\, g^{rs}{}_{,0}\, dx^0 - \left\{\int \mathfrak{J}^s\, dx^0\right\}_{;s} = 0. \end{cases}$$

These are the identities of the field equations (35), (36).

The theory involves four universal constants, $\underset{2}{\alpha}/\underset{1}{\alpha}$, $\underset{3}{\alpha}/\underset{1}{\alpha}$, $\underset{4}{\alpha}/\underset{1}{\alpha}$, λ. The constant $\underset{2}{\alpha}/\underset{1}{\alpha}$ corresponds to the gravitational constant involving a connection between the (arbitrary) units of length and mass. One of the other constants, for instance λ, will depend on the unit of length. The remaining two are "genuine" universal constants which cannot be eliminated from the theory.

SUMMARY

Kaluza's five dimensional theory of the physical space provides a unitary representation of gravitation and electromagnetism. It demands, if consistently interpreted as a five dimensional theory, the existence of a vector A, whose symmetrical covariant derivatives vanish in the whole space. Apart from its physical meaning, this assumption seems to be artificial. Furthermore the representation of a purely four dimensional continuum by a five dimensional one appears not to be sufficiently justified.

By our generalization we try to eliminate these objections. The hypothesis of cylindricity, (i.e. of the existence of a vector A with vanishing symmetrical derivatives) is replaced by the assumption that the space is closed in the direction of the fifth coördinate. By this change the basic assumptions of the theory are considerably simplified. Furthermore it is much more satisfactory to introduce the fifth dimension not only formally, but to assign to it some physical meaning. Nevertheless there is no contradiction with the empirical four dimensional character of physical space.

APPENDIX

Analysis of Tensor Densities

The tensor density plays, beside the concept of tensors, an important rôle in relativistic investigations. Although the concept of tensor densities can easily be reduced to that of tensors nevertheless the whole formalism and especially the derivation of the field equations from the variational principle are considerably simplified if tensor densities are introduced independently. Although the analytical properties of tensor densities have been fully investigated[7] we shall, in order to make the reading easier, make some remarks on the subject.

A tensor density differs from a tensor by a factor (common to all components) having the transformation character of a scalar density. A scalar density $\underset{(n)}{\rho}$ of the "weight" n is a quantity with the following transformation property:

$$\underset{(n)}{\bar{\rho}} = \underset{(n)}{\rho} \left| \frac{\partial x^{\alpha}}{\partial \bar{x}^{\beta}} \right|^{n}. \tag{A 1}$$

There exist, correspondingly, tensor densities of different weights. (Ordinary tensors are tensor densities of the weight zero.) By multiplying two tensor densities of the weights m and n we obtain a tensor density of the weight $(m + n)$. The rules of a tensor density algebra follow immediately from that of the tensor algebra and need not be derived here.

For the sake of simplicity we shall consider here only quantities in a four dimensional continuum; the results are, however, independent of the number of dimensions.

There exist, in every four dimensional space, the two tensor densities of Levi-Civita

$$\underset{(-1)}{\delta_{iklm}}$$

and

$$\underset{(+1)}{\delta_{iklm}},$$

antisymmetric in all indices, the components of which are $+1$ or -1, depending of whether $(iklm)$ is an odd or an even permutation of (1234). Their tensor density character follows immediately from the transformation law of tensor densities. From the Levi Civita tensor densities the well known Kronecker tensor $\delta_m^{m'}$ (with components equal 1 if $m = m'$ and 0 if $m \neq m'$) can be obtained by:

$$\delta_m^{m'} = \tfrac{1}{6} \underset{(-1)}{\delta_{iklm}} \underset{(+1)}{\delta^{iklm'}}. \tag{A 2}$$

We assume in the space under consideration the existence of symmetrical components of parallel displacement (the space functions $\Gamma^l_{ik} = \Gamma^l_{ki}$) by which the covariant derivatives of a tensor are defined in the usual way:

$$B^{k\cdots}_{i\cdots;s} = B^{k\cdots}_{i\cdots,s} + B^{l\cdots}_{i\cdots}\Gamma^k_{ls} + \cdots - B^{k\cdots}_{l\cdots}\Gamma^l_{is} - \cdots . \tag{A 3}$$

This definition of the covariant differentiation has two essential properties:

1. It satisfies the rule for the differentiation of products of the ordinary calculus:

$$(AB)_{;s} = AB_{;s} + BA_{;s}.$$

---

[7] V. Hlavatý, Annalii di Mathematica, V, Ser. IV (1927/28): *Théorie des densités dans le displacement général.*

2. The absolute derivatives of the Kronecker tensor vanish identically:

$$\delta^k_{i;s} = 0\,.$$

We shall now define the covariant derivative of a tensor density so that:

1. The rule for the differentiation of products is preserved;
2. The absolute derivatives of the co- and contravariant tensor densities of Levi Civita vanish identically.

By direct calculation we confirm that these demands are satisfied by the following definition:

$$\underset{(n)}{B}{}^{k\cdots}_{i\cdots;s} = \underset{(n)}{B}{}^{k\cdots}_{i\cdots,s} + \underset{(n)}{B}{}^{l\cdots}_{i\cdots}\Gamma^k_{ls} + \cdots - \underset{(n)}{B}{}^{k\cdots}_{l\cdots}\Gamma^l_{is} - \cdots - n\,\underset{(n)}{B}{}^{k\cdots}_{i\cdots}\Gamma^r_{sr}\,. \tag{A 4}$$

This is the rule for the differentiation of tensor densities of the weight n.

If we specialize the space further by assuming that it is Riemannian then the components of the parallel displacement are given by:

$$\Gamma^s_{mn} = \tfrac{1}{2}g^{st}(g_{ml,n} + g_{nl,m} - g_{mn,l}), \tag{A 5}$$

and therefore

$$\Gamma^n_{sn} = \frac{1}{2}\frac{g_{,s}}{g}, \qquad g = |g_{ik}|. \tag{A 6}$$

The determinant g is a scalar density of the weight 2. By applying (A 4) we get:

$$\underset{(2)}{g}{}_{;s} = g_{,s} - 2g\left(\frac{1}{2}\frac{g_{,s}}{g}\right) = 0. \tag{A 7}$$

Therefore the absolute derivative of $\sqrt{g}$ and, because of the rule for the differentiation of products, the absolute derivatives of $g_{mn}\sqrt{g}$ and of $g^{mn}\sqrt{g}$ vanish also.

We shall now show how the tensor densities can be applied in the variational process.

1. The derivation of the equations of gravitation by the variation of Riemann's curvature tensor:

$$\left\{\begin{aligned} &\delta\int \underset{(1)}{\mathfrak{H}}\,d\tau = 0,\\ &\underset{(1)}{\mathfrak{H}} = \sqrt{-g}\,g^{km}\,\delta^l_i\,R^i_{.klm}\,. \end{aligned}\right. \tag{A 8}$$

It was shown by Palatini that the variation with respect to the g^{km} can be simplified, because of the equation:

$$\delta R^i_{.klm} = (\delta\Gamma^i_{kl})_{;m} - (\delta\Gamma^i_{km})_{;l} \tag{A 9}$$

which can easily be verified (note that $\delta\Gamma^i_{kl}$ is a tensor!). Because of

$$\delta(\sqrt{-g}\,g^{km}) = \sqrt{-g}(\delta g^{km} - \tfrac{1}{2}g^{km}g_{ab}\cdot\delta g^{ab}), \tag{A 10a}$$

we have then

$$\delta \int \underset{(1)}{\mathfrak{H}}\, d\tau = \int \sqrt{-g}(R_{km} - \tfrac{1}{2}g_{km}R)\delta g^{km}\, d\tau \tag{A 10b}$$
$$+ \int \sqrt{-g}\, g^{km}[(\delta\Gamma^l_{kl})_{;m} - (\delta\Gamma^l_{km})_{;l}]\, d\tau.$$

As $(g^{km}\cdot\sqrt{-g})_{;l}$ vanishes the integrand of the second integral may be written

$$[\sqrt{-g}(g^{kl}\,\delta\Gamma^m_{km} - g^{km}\,\delta\Gamma^l_{km})]_{;l}, \tag{A 10c}$$

where the square bracket expression $\underset{(1)}{K}{}^l$ is a vector density of the weight 1. It follows, therefore, from (A 4) that $\underset{(1)}{K}{}^l_{;l}$ may be replaced by $\underset{(1)}{K}{}^l_{,l}$. This integral vanishes, according to Gauss' theorem, if the variations of the g_{ik} and of the Γ^l_{ik} vanish on the boundary of the integration region. There remains only the first integral of (A 10b) so that the result is

$$R_{km} - \tfrac{1}{2}g_{km}R = 0. \tag{A10}$$

2. Tensor densities in the space of the generalized Kaluza theory. In such a space, and in the special coördinate system, a scalar density of the weight n can be defined by the transformation law:

$$\underset{(n)}{\bar{\rho}} = \underset{(n)}{\rho}\left|\frac{\partial x^\alpha}{\partial \bar{x}^\beta}\right|^n \quad \text{for four-transformations,} \tag{A 11}$$
$$\underset{(n)}{\bar{\rho}} = \underset{(n)}{\rho} \quad \text{for cut-transformations.}$$

The transformation law of tensor densities is also determined by this definition, i.e. reduced to that of tensors.

Covariant differentiation of tensor densities: We can expect, because of (A 4) and (27), (28), that the following formula expresses the differentiation rule:

$$\underset{(n)}{B}{}^{k\cdots}_{i\cdots;s} = \underset{(n)}{B}{}^{k\cdots}_{i\cdots,s} - A_s \underset{(n)}{B}{}^{k\cdots}_{i\cdots,0} + \underset{(n)}{B}{}^{l\cdots}_{i\cdots}\Gamma^k_{ls} \tag{A 12}$$
$$+ \cdots - \underset{(n)}{B}{}^{k\cdots}_{l\cdots}\Gamma^l_{is} - \cdots - n\, B^{k\cdots}_{i\cdots}\Gamma^l_{sl}.$$

$\underset{(-1)}{\delta}{}_{iklm}$ is a tensor density with respect to four-transformations and invariant with respect to cut-transformations. It is, therefore, a tensor density according to our definitions. The absolute derivative of this tensor density, given by (A 12), vanishes. It can also be seen that (A 12) preserves the rule for the differentation of products.

It follows from our definitions that $\underset{(2)}{g} = |g_{mn}|$ is a scalar density of the weight 2 with the absolute derivatives:

$$\underset{(2)}{g}{}_{;s} = g_{,s} - A_s g_{,0} - 2g\Gamma^l_{sl} = 0. \tag{A 13}$$

(This vanishes because of (27b).) Furthermore the equation holds:

(A 14) $$g_{mn;s} = 0.$$

We form the divergence $\underset{(1)}{B}{}^{s}_{;s}$ of a vector density $\underset{(1)}{B}{}^{s}$ of the weight 1. According to (A 12) we have

(A 15a) $$\underset{(1)}{B}{}^{s}_{;s} = \underset{(1)}{B}{}^{s}_{,s} - A_s \underset{(1)}{B}{}^{s}{}_{,0},$$

and as A_s does not depend on x^0

(A 15) $$\underset{(1)}{B}{}^{s}_{;s} = \underset{(1)}{B}{}^{s}_{,s} - (A_s \underset{(1)}{B}{}^{s})_{,0}.$$

It is essential to note here that both terms on the right hand side take the form of ordinary derivatives.

We have now, as in the gravitational theory, to form the variation of the integral (Compare (32), (33))

(A 16a) $$\int \sqrt{-g}\, g^{km} \delta^l_i R^i_{.klm}\, d\tau.$$

This integral, however, is to be taken over a region of the variables $x^0, x^1 \cdots x^4$, which contains exactly one period of the coördinate x^0 (and is arbitrary with respect to $x^1 \cdots x^4$). The $R^i_{.klm}$ are defined by (30a). The variation of this quantity takes the form:

(A 16b) $$\delta R^i_{.klm} = (\delta\Gamma^i_{kl})_{,m} - (\delta\Gamma^i_{km})_{;l} - \delta A_m \Gamma^i_{kl,0} + \delta A_l \Gamma^i_{km,0}.$$

The variation of this integral gives, as far as the variation of the factor $g^{km} \cdot \sqrt{-g}$ is concerned, the contribution:

(A 16c) $$\int \sqrt{-g}(\tfrac{1}{2}R_{km} + \tfrac{1}{2}R_{mk} - \tfrac{1}{2}g_{km}R)\delta g^{km}\, d\tau.$$

The variation of $R^i_{.klm}$ gives, because of the two last terms in (A 16b),

(A 16d) $$\int \sqrt{-g}(g^{km}\Gamma^l_{km,0} - g^{kl}\,\Gamma^m_{km,0})\delta A_l\, d\tau.$$

The first two terms on the right hand side of (A 16b), however, do not give any contribution. The proof is similar for both terms; we shall, therefore, show it only for the first one. The integrand

$$\sqrt{-g} \cdot g^{km}(\delta\Gamma^l_{kl})_{;m}$$

can, because of

$$(\sqrt{-g} \cdot g^{km})_{;m} = 0,$$

(following from (A 13), (A 14)) be brought into the form:

$$\underset{(1)}{B}{}^{m}_{;m}.$$

But such an integral vanishes because of (A 15). It does so partially because the variations vanish on the x^a-boundaries, partially because of the periodicity of B^m with respect to x^0. The final result of the variation is:

$$\text{(A 16)} \quad \int \sqrt{-g}\{(\tfrac{1}{2}R_{km} + \tfrac{1}{2}R_{mk} - \tfrac{1}{2}g_{km}R)\delta g^{km} + (g^{km}\Gamma^l_{km,0} - g^{kl}\Gamma^m_{km,0})\delta A_l\}\, d\tau,$$

where the δg^{km} depend arbitrarily on x^a, x^0, the δA_s however only on x^a.

The calculations applied in the derivation of the identities are quite similar.

INSTITUTE FOR ADVANCED STUDY.

GEOMETRY. - The Equations of Kaluza's Unified Theory

Note[1] by **M. Yves Thiry,** presented by M. Elie Cartan

Kaluza's[2] attempt at a unified theory consists of considering space-time as the $x^0 =$ const. subspace of a five-dimensional Riemann space, and of assuming this subspace cylindrical with respect to the fifth coordinate x^0. We shall show here that the 15 rigorous equations of this theory can be established by a very simple method when using a *répère mobile* and the techniques of exterior calculus.[3] In this paper greek indices vary from 0 to 4, latin indices from 1 to 4; barred quantities belong to space-time; $\partial_i f$ is the coefficient of the differential from ω_i in df; the symbol Σ is omitted when summing over repeated indices; $\overline{\nabla}_i$ is the covariant derivative and $\overline{\Delta}_2$ the laplacian.

Let $d\sigma^2 = \gamma_{\lambda\mu} dx^\lambda dx^\mu$ be the metric of the five-dimensional space. Set $\gamma_{00} = V^2$ and write $d\sigma^2$ as a sum of squares in the form

$$d\sigma^2 = \sum_\alpha (\omega_\alpha)^2$$

with

$$\omega_0 = V(dx^0 + \beta\phi_i dx^i) = V(dx^0 + \beta\Phi_i\omega_i), \quad \sum_i(\omega_i)^2 = ds^2 = g_{ij}dx^i dx^j,$$

or

$$g_{ij} = \gamma_{ij} - \beta^2 V^2 \phi_i\phi_j, \quad \phi_i = \frac{\gamma_{io}}{\beta V^2} \ \ (\beta = const.),$$

ϕ_i and Φ_i being the components of the same vector in the natural and the mobile orthonormal frame respectively.

Let us first calculate the coefficients $\gamma_{\alpha\beta\lambda}$ and $c_{\lambda\mu\alpha}$ defined by

$$\omega_{\alpha\beta} = \gamma_{\alpha\beta\lambda}\omega_\lambda; \ \ d\omega_\alpha = \frac{1}{2} c_{\lambda\mu\alpha}\omega_\lambda \wedge \omega_\mu \ .$$

---

[1] Meeting of 12 January 1948.

[2] Sitzungsber. d. Preuss, Akad. d. Wiss., 1921, p. 966.

[3] E. Cartan, Leçons sur la théorie des espaces de Riemann, Paris, 1946, p. 221; Lichnerowicz, Algèbre linéaire, Paris, 1947, p. 143-172.

One finds

$$c_{hki} = c_{hki}; \quad \bar{\gamma}_{hki} = \gamma_{hki};$$

$$c_{oki} = 0; \quad c_{i00} = \gamma_{i00} = \frac{\partial_i V}{V}\,;$$

$$c_{ik0} = 2\gamma_{ik0} = 2\gamma_{i0k} = \beta V F_{ik} \qquad (\text{or } F_{ik} = \bar{\nabla}_i \Phi_k - \bar{\nabla}_k \Phi_i).$$

One can then calculate the components of the Riemann-Christoffel tensor from the formula

$$d\omega_{\alpha\beta} = \omega_{\alpha\lambda} \wedge \omega_{\lambda\beta} + \frac{1}{2}\, R_{\alpha\beta\lambda\mu} \omega_\lambda \wedge \omega_\mu,$$

and deduce therefrom the components of the contracted tensor

$$R_{ik} = \bar{R}_{ik} + \frac{1}{V} \bar{\nabla}_k (\partial_i V) - \frac{\beta^2 V^2}{2} F_{ij} F_{jk},$$

$$R_{io} = \frac{\beta V}{2} \bar{\nabla}_j F_{ji} + \frac{3\beta}{2} \partial_j V F_{ji},$$

$$R_{00} = \frac{1}{V} \bar{\Delta}_2 V - \frac{\beta^2 V^2}{4}\, \mathbf{H}^2 \qquad \left[\text{or } \mathbf{H}^2 = \sum_{i,j} (F_{ij})^2 \right].$$

One also has

$$\bar{R} = R - \frac{2}{V} \bar{\Delta}_2 V - \frac{\beta^2 V^2}{4.} \mathbf{H}^2,$$

and the identity

$$\bar{\nabla}_k F_{kl} + \bar{\nabla}_k F_{li} + \bar{\nabla}_l F_{ik} = 0. \tag{1}$$

The first fourteen field equations $S_{\alpha\beta} = [R_{\alpha\beta} - \frac{1}{2} \delta_{\alpha\beta} R] = 0$ (α,β not both zero) can then be written as

$$\bar{S}_{ik} = \frac{\beta^2 V^2}{2} \left(\frac{1}{4} \delta_{ik} \mathbf{H}^2 - F_{ij} F_{kj}\right) - \frac{1}{V} [\bar{\nabla}_k (\partial_i V) - \bar{\Delta}_2 V], \tag{2}$$

$$\bar{\nabla}_j F_{ji} = -\, 3 \frac{\partial_j V}{V} F_{ji}. \tag{3}$$

As to the fifteenth equation, it can be written either as $R_{00} = 0$, or in other words,

$$\frac{1}{V}\overline{\Delta}_2 V - \frac{\beta^2 V^2}{4}\mathbf{H}^2 = 0, \tag{4}$$

or in the form $R^0_0 = 0$, by calculating this component of the contracted tensor with respect to the natural frame, with the result

$$\frac{1}{V}\overline{\nabla}_i h_i = 0 \ \text{ with } \ h_i = \partial_i V - \frac{\beta^2 V^3}{2}\Phi_i F_{ji}. \tag{5}$$

The equations (2) generalise Einstein's equations with a source term, the equations (1) and (3) furnish two groups of equations analogous to those of Maxwell-Lorentz; we shall return at some later time to the physical meaning that one can propose for the fifteenth equation in the form (4) or (5). We wish to point out however that in the case $V =$ const. [O. Klein's theory [4]], equation (4) yields $\mathbf{H}^2 = 0$, which, given the signature of ds^2, would require the square of the electric field to equal the square of the magnetic field. Moreover the equations (4) and (5) applied to an exterior static compact space-time, correspond to the two equations obtained by Lichnerowicz[5], and allow one to study the regularity of such a space-time.

Translation from the French original by P.G.O.F.

GÉOMÉTRIE. — *Les équations de la théorie unitaire de Kaluza.*

Note (1) de M. **Yves Thiry**, présentée par M. Élie Cartan.

La tentative de théorie unitaire de Kaluza (2) consiste à considérer l'espace-temps comme un sous-espace $x^0 = \text{const.}$ d'un espace de Riemann à cinq dimensions, en supposant ce sous-espace cylindrique par rapport à la cinquième coordonnée x^0. Nous nous proposons ici de montrer que les 15 équations rigoureuses de cette théorie peuvent être établies par une méthode très simple, utilisant un repère mobile et la technique du calcul extérieur (3). Dans cet

---

(1) Séance du 12 janvier 1948.

(2) *Sitzungsber. d. Preuss. Akad. d. Wiss.*, 1921, p. 966.

(3) E. Cartan, *Leçons sur la théorie des espaces de Riemann*, Paris, 1946, p. 221; Lichnerowicz, *Algèbre linéaire*, Paris, 1947, p. 143-172.

exposé, les indices grecs varient de 0 à 4, les indices latins de 1 à 4; les éléments surlignés appartiennent à l'espace-temps; $\partial_i f$ est le coefficient de la forme différentielle ω_i dans df; le signe Σ est omis quand il faut sommer par rapport à un même indice répété deux fois; $\overline{\nabla}_i$ désigne l'opérateur de dérivation covariante, $\overline{\Delta}_2$ le laplacien.

Soit $d\sigma^2 = \gamma_{\lambda\mu}\, dx^\lambda\, dx^\mu$ la métrique de l'espace à cinq dimensions. Posons $\gamma_{00} = V^2$ et décomposons $d\sigma^2$ en carrés sous la forme

$$d\sigma^2 = \sum_\alpha (\omega_\alpha)^2$$

avec

$$\omega_0 = V(dx^0 + \beta\varphi_i\, dx^i) = V(dx^0 + \beta\Phi_i\omega_i), \qquad \sum(\omega_i)^2 = ds^2 = g_{ij}\, dx^i\, dx^j,$$

où

$$g_{ij} = \gamma_{ij} - \beta^2 V^2 \varphi_i\varphi_j, \qquad \varphi_i = \frac{\gamma_{i0}}{\beta V^2} \qquad (\beta = \text{const.}),$$

φ_i et Φ_i étant les composantes d'un même vecteur dans le repère naturel et dans le repère mobile orthonormé.

Calculons d'abord les coefficients $\gamma_{\alpha\beta\lambda}$ et $c_{\lambda\mu\alpha}$ définis par

$$\omega_{\alpha\beta} = \gamma_{\alpha\beta\lambda}\,\omega_\lambda; \qquad d\omega_\alpha = \frac{1}{2} c_{\lambda\mu\alpha}\,\omega_\lambda \wedge \omega_\mu.$$

On trouve

$$\begin{aligned} &c_{hki} = c_{hki}; && \bar{\gamma}_{hki} = \gamma_{hki}; \\ &c_{0ki} = 0; && c_{i00} = \gamma_{i00} = \frac{\partial_i V}{V}; \\ &c_{ik0} = 2\gamma_{ik0} = 2\gamma_{i0k} = \beta V F_{ik} && (\text{où } F_{ik} = \overline{\nabla}_i\Phi_k - \overline{\nabla}_k\Phi_i) \end{aligned}$$

On peut ensuite calculer les composantes du tenseur de Riemann-Cristoffel en utilisant la formule

$$d\omega_{\alpha\beta} = \omega_{\alpha\lambda} \wedge \omega_{\lambda\beta} + \frac{1}{2} R_{\alpha\beta\lambda\mu}\,\omega_\lambda \wedge \omega_\mu,$$

et en déduire les composantes suivantes du tenseur contracté

$$\begin{aligned} R_{ik} &= \overline{R}_{ik} + \frac{1}{V}\overline{\nabla}_k(\partial_i V) - \frac{\beta^2 V^2}{2} F_{ij}F_{jk}, \\ R_{i0} &= \frac{\beta V}{2}\overline{\nabla}_j F_{ji} + \frac{3\beta}{2}\partial_j V\, F_{ji}, \\ R_{00} &= \frac{1}{V}\overline{\Delta}_2 V - \frac{\beta^2 V^2}{4} H^2 \qquad \left[\text{où } H^2 = \sum_{i,j} (F_{ij})^2\right]. \end{aligned}$$

On a de plus

$$\overline{R} = R - \frac{2}{V}\overline{\Delta}_2 V - \frac{\beta^2 V^2}{4} H^2,$$

et l'identité

$$\overline{\nabla}_i F_{kl} + \overline{\nabla}_k F_{li} + \overline{\nabla}_l F_{ik} = 0. \tag{1}$$

Les quatorze premières équations du champ $S_{\alpha\beta} = [R_{\alpha\beta} - (1/2)\delta_{\alpha\beta} R] = 0$ (α, β non tous deux nuls) s'écrivent alors

$$\overline{S}_{ik} = \frac{\beta^2 V^2}{2}\left(\frac{1}{4}\delta_{ik} H^2 - F_{ij} F_{kj}\right) - \frac{1}{V}\left[\overline{\nabla}_k(\partial_i V) - \delta_{ik}\overline{\Delta}_2 V\right], \tag{2}$$

$$\overline{\nabla}_j F_{ji} = -3\frac{\partial_j V}{V} F_{ji}. \tag{3}$$

Quant à la quinzième équation, elle peut s'écrire soit sous la forme $R_{00} = 0$, c'est-à-dire

$$\frac{1}{V}\overline{\Delta}_2 V - \frac{\beta^2 V^2}{4} H^2 = 0, \tag{4}$$

soit sous la forme $R_0^0 = 0$ en calculant cette composante du tenseur contracté relative au repère naturel, ce qui donne :

$$\frac{1}{V}\overline{\nabla}_i h_i = 0 \qquad \text{avec} \quad h_i = \partial_i V - \frac{\beta^2 V^3}{2}\Phi_i F_{ji}. \tag{5}$$

Les équations (2) généralisent les équations d'Einstein avec second membre; les équations (1) et (3) fournissent deux groupes d'équations analogues à celles de Maxwell-Lorentz; nous reviendrons ultérieurement sur la signification physique que l'on peut proposer pour la quinzième équation sous les formes (4) ou (5). Signalons cependant que dans le cas V = const. [théorie d'O. Klein (4)], l'équation (4) donne $H^2 = 0$, ce qui, en tenant compte de la signature du ds^2, conduirait à égaler le carré du champ électrique au carré du champ magnétique. De plus les équations (4) et (5), appliquées à un espace-temps extérieur statique compact, correspondent à deux équations établies par Lichnerowicz (5) et permettent par suite l'étude de la régularité d'un tel espace-temps.

---

(4) *Zeitschrift für Physik*, **37**, 1926, p. 895.

(5) *Comptes rendus*, **222**, 1946, p. 432, équations (2′) et (1′).

CHAPTER 3

Non-Abelian Generalization

Reprinted from: Dynamical Theory of Groups and Fields,
1963, Les Houches, Lecture Notes by B.S. DeWitt.

Problem 77. Show that the dynamical theory of the combined gravitational and Yang–Mills fields may be given the following purely geometrical interpretation: (1) Increase the dimensionality of space-time from 4 to $4 + m$. (2) Introduce coordinates x^{μ}, x^{α} and a metric tensor of the form

$$\begin{pmatrix} \mathbf{g}_{\alpha\beta} & \mathbf{g}_{\alpha\nu} \\ \mathbf{g}_{\mu\beta} & \mathbf{g}_{\mu\nu} \end{pmatrix}$$

where $\mathbf{g}_{\alpha\beta}$ depends only on the x^{α} and is the natural metric of the generating group manifold in an arbitrary fixed coordinate system. (3) Hold the functional form of $\mathbf{g}_{\alpha\beta}$ rigid by allowing, as invariance transformations, only general transformations of the x^{μ} and x^{μ}-dependent isometric transformations of the x^{α} generated by the Yang–Mills group. (4) Form the Riemann scalar determined by the above metric, and multiply it by the square root of the negative of the determinant of the metric. (5) Introduce as action functional, with the $\mathbf{g}_{\alpha\mu}$ and the $\mathbf{g}_{\mu\nu}$ as dynamical variables, the integral of this density over the x^{μ} and the x^{α}. (6) By rewriting the metric in the form

$$\begin{pmatrix} \mathbf{g}_{\alpha\beta} & \mathbf{g}_{\alpha\nu} \\ \mathbf{g}_{\mu\beta} & \mathbf{g}_{\mu\nu} \end{pmatrix} = \begin{pmatrix} \gamma_{\gamma\delta} L^{-1\gamma}{}_{\alpha} L^{-1\delta}{}_{\beta} & \gamma_{\gamma\delta} L^{-1\gamma}{}_{\alpha} A^{\delta}{}_{\nu} \\ \gamma_{\gamma\delta} A^{\gamma}{}_{\mu} L^{-1\delta}{}_{\beta} & g_{\mu\nu} + \gamma_{\gamma\delta} A^{\gamma}{}_{\mu} A^{\delta}{}_{\nu} \end{pmatrix}$$

where the auxiliary functions $L^{-1\alpha}{}_{\beta}$ depend only on the x^{α} while the $A^{\alpha}{}_{\mu}$ and $g_{\mu\nu}$ depend only on the x^{μ}, and by independently varying the $A^{\alpha}{}_{\mu}$ and the $g_{\mu\nu}$, show that this action leads to precisely the dynamical equations (16.60) and (16.61) previously obtained, minus the field φ.

This result forms the basis for the existence of a class of so-called *unified field theories*,* and suggests that geometry should perhaps provide the foundation for all of physics. If this is granted, however, there is then no reason to hold the geometry of the cross-sections, $x^{\mu} =$ constant, rigid. The $\mathbf{g}_{\alpha\beta}$ themselves should be subject to dynamical variation, and the Yang–Mills group should express only an approximate symmetry of the system. The lack of direct tactile evidence for the extra dimensions of space-time could be regarded as due to extreme smallness of the average volume of the cross sections $x^{\mu} =$ constant. A possible advantage of such a model would be a restriction on the types of representations which could be selected for any additional fields φ which might be introduced. The transformation character of such fields in ordinary space-time would have to be related to their transformation properties under the Yang–Mills group. In this connection the *topology* selected for the cross sections $x^{\mu} =$ constant would be of fundamental importance.

Generalization of the Kaluza-Klein theory for an arbitrary non-abelian gauge group

by

Ryszard KERNER
Institute of Theoretical Physics, University of Warsaw,
Warsaw, Poland.

RÉSUMÉ. — L'isomorphisme entre les théories d'électromagnétisme d'Utiyama et de Kaluza et Klein est généralisé pour un groupe non-abelien semi-simple de jauge. Les équations de champ de jauge non-abelien sont dérivées d'un principe variationnel géométrique. L'interaction entre le champ gravitationnel et le champ généralisé de Yang et Mills est donnée sous une forme explicite. Les équations de mouvement d'une particule dans le champ extérieur sont obtenus à partir de l'équation des géodésiques dans l'espace fibré principal.

SUMMARY. — The isomorphism between the Utiyama and the Kaluza-Klein theories of electromagnetism is generalized to the case of a non-abelian semi-simple gauge group. The equations of the non-abelian gauge field are derived from a geometrical least action principle. The interaction between the gravitational and the generalized Yang-Mills field is given explicitly. The equations of motion of a particle in an external field are obtained from the geodetic equation in the principal fibre bundle space.

1. THE EQUIVALENCE OF THE UTIYAMA AND KALUZA-KLEIN APPROACHES TO THE THEORY OF THE ELECTROMAGNETIC FIELD

Utiyama has put forward a theory of invariant interaction defining the quantities on which the lagrangian invariant with respect to a given gauge group must depend [5]. In the case of electromagnetic field this theory can be regarded as a theory of connections in a principal fibre bundle over a Riemannian manifold V_4 with a structure group isomorphic with a one-dimensional sphere S^1. In this case there exists an isomorphism between the theory of Utiyama and the theory of Kaluza and Klein that treats the electromagnetic and gravitational field in a unified way, as some special metric geometry in a five-dimensional differential manifold. The existence of such an isomorphism was shown recently by Trautman and Tulczyjew [*6*], [*7*].

The possibility of introducing a metric in the principal fibre bundle, which in this case is also a five-dimensional manifold, is inherent in the very structure of a bundle with connection. This special metric is defined by the following requirements:

a) The horizontal subspaces of the tangent space to the bundle must be orthogonal in this metric to the vertical subspaces;

b) The projection of the metric onto the horizontal space must be isomorphic with the Riemannian metric of the base manifold;

c) The vertical part of the metric must be isomorphic to some metric of the space tangent to the fibre, *i. e.* to some metric on the Lie algebra of the structure group.

The only arbitrariness present in the definition of the metric is contained in *c)*; indeed, there are many different metrics on the group. For the Kaluza-Klein theory the group metric is the most trivial, defined by the notion of the natural length in R^1. This choice is equivalent to the « strong cylindricity condition » of the Kaluza-Klein theory (cf. Einstein [*2*]).

Utiyama's theory does not define the lagrangian; it only dictates its dependence on geometric quantities, *i. e.* on the curvature form of the bundle. However the explicit form of this dependence is not defined. In order to generalize the equivalence of the Utiyama and Kaluza-Klein approaches for non-abelian gauge groups, we shall assume that the lagrangian is equal to the scalar density made of the square root of the determinant of the Riemannian metric tensor multiplied by the scalar Riemannian curvature, R, of the bundle manifold. There still remains the problem of choosing a metric on the group space. We shall assume the metric group invariant, and for the case of a semi-simple Lie group we shall take the Killing tensor constructed from the structure constants of the group:

$$g_{ab} = C^c_{ad} C^d_{cb}$$

It is easy to see, that for a one-dimensional abelian gauge group (*i. e.* in the theory of gravitation and electromagnetism) the Riemann scalar of the bundle, constructed with the metric tensor defined above, is equal to the sum of the Riemann scalar of the base manifold and the lagrangian of the electromagnetic field, and that the determinant of the five-dimensional metric tensor is equal to the determinant of the metric tensor of the base. However, it is by no means obvious that for any non-abelian gauge group the Riemannian scalar curvature of the bundle will be the sum of the Riemann scalar of the base and a lagrangian of the Yang-Mills type. Our first aim is to prove this statement.

2. GENERALIZATION OF THE KALUZA-KLEIN THEORY FOR A NON-ABELIAN GAUGE GROUP

Following the program given above, we construct a principal fibre bundle over some Riemannian space-time manifold V_4 with metric tensor g_{ij}, the structure group being some finite-dimensional Lie group G. We assume that there is a connection in the bundle, given by a Lie algebra valued 1-form A on the bundle manifold. The curvature form of this connection, being the covariant differential of the connection form, DA, is interpreted as a generalized Yang-Mills field tensor. We also assume an invariant metric on the structure group, choosing the Killing form in the case of a semi-simple group.

In the following, we restrict our considerations to a neighbourhood of some point of the bundle; it is well known that such a neighbourhood can be regarded as a direct product of some neighbourhood in the base and a neighbourhood in the group space. This suggests choosing special coordinate system natural to the problem. Let the indices $i, j, k, \ldots$ take values from 1 to 4, the indices $a, b, c, \ldots$ – from 4 to M + 4, M = dim G, and the indices $\alpha, \beta, \gamma, \ldots$ – from 1 to M + 4. The natural choice is taking the first 4 of them as the coordinates in the base neighbourhood, and the remaining M as the coordinates in the neighbourhood of the group. When this is done, the differential of the projection onto the base, dp, has the coordinates $dp^i_j = \delta^i_j$ and $dp^i_a = 0$. Similarly, the isomorphism σ of the Lie algebra of the structure group onto the vertical subspace of the tangent space to the bundle has the coordinates

$$\sigma^a_b = \delta^a_b, \qquad \sigma^i_b = 0.$$

The connection form A has coordinates A^a_j, and the curvature form in our coordinate frame is

$$F^a_{ij} = \frac{1}{2}\left(\partial_i A^a_j - \partial_j A^a_i\right) + C^a_{bc} A^b_i A^c_j \tag{1}$$

C^a_{bc} being the structure constants of the group G. Other components of the curvature form vanish in our coordinate frame.

The coordinate system in the base neighbourhood can be chosen at will, whereas in the group neighbourhood, in order to simplify the calculations, we choose locally geodesic coordinates. That is, we require the covariant derivative of A^a_i in the group space to vanish; as the connection coefficients in the group space are equal to the structure constants, this means that

$$\partial_a A^b_j + C^b_{ac} A = 0 \tag{2}$$

or

$$\partial_a A^b_j = C^b_{ca} A^c_j \tag{3}$$

because of the antisymmetry of C^a_{bc} in the two lower indices.

Let the metric of the base be g_{ij} and the invariant metric on the group be g_{ab}. Here invariance means that

$$R_a g = g = L_a g \quad , \quad a \in G \tag{4}$$

where R_a and L_a are the right and left translations respectively. Without lack of generality we can reduce (4) to the condition

$$\partial_a g_{bc} = 0 \tag{5}$$

Local triviality of the bundle gives

$$\partial_a g_{ij} = 0 \qquad\qquad \partial_j g_{ab} = 0 \tag{6}$$

Let us now find the explicit form of the bundle metric satisfying the requirements *a), b), c)*. Expressed in coordinates these requirements are

$$dp^i_\alpha dp^j_\beta g^{\alpha\beta} = g^{ij}, \qquad\qquad A^a_\alpha A^b_\beta g^{\alpha\beta} = g^{ab} \tag{7}$$

Analogous equations can be written for the covariant tensor:

$$d\tau^\alpha_i d\tau^\beta_j g_{\alpha\beta} = g_{ij}, \qquad\qquad \sigma^\alpha_a \sigma^\beta_b g^{\alpha\beta} = g_{ab} \tag{8}$$

where $d\tau^\alpha_i$ is the differential of the lift and σ^α_a is the isomorphism of the

Lie algebra onto the vertical subspace tangent to the fibre. By virtue of the general relations in principal fibre bundles

$$A^a_\alpha \sigma^\alpha_b = \delta^a_b, \qquad dp^i_\alpha d\tau^\alpha_j = \delta^i_j,$$
$$A^a_\alpha d\tau^\alpha_j = 0, \qquad dp^i_\alpha \sigma^\alpha_a = 0. \tag{9}$$

one can easily deduce that the only solution for $g_{\alpha\beta}$ and $g^{\alpha\beta}$ satisfying conditions (7) and (8) is:

$$g_{\alpha\beta} = dp^i_\alpha dp^j_\beta g_{ij} + A^a_\alpha A^b_\beta g_{ab} \tag{10}$$
$$g^{\alpha\beta} = d\tau^\alpha_i d\tau^\beta_j g^{ij} + \sigma^\alpha_a \sigma^\beta_b g^{ab} \tag{11}$$

In our special coordinates these formulae take the form:

$$g_{\alpha\beta} = \left(\begin{array}{c:c} g_{ij} + g_{ab}A^a_i A^b_i & g_{ab}A^\alpha_i \\ \hdashline g_{ab}A^b_j & g_{ab} \end{array}\right) \tag{12}$$

and

$$g^{\alpha\beta} = \left(\begin{array}{c:c} g^{ij} & -g^{ij}A^b_j \\ \hdashline -g_{ij}A^a_i & g^{ab} + g^{ij}A^a_i A^b_j \end{array}\right) \tag{13}$$

A simple computation shows that (13) is the inverse of (12) and that the determinant of (12) is equal to the determinant of g_{ij}. With these preparations we can calculate the Christoffel symbols for the bundle manifold. The calculations are done in the coordinate system defined above. It is useful to remark, that the field of vectors $A^\alpha_a = g^{\alpha\beta}g_{ab}A^b_\beta$ is a Killing field with respect to our metric. This can be easily seen in our coordinate system. First, one can easily see that the only non vanishing components of A^α_a are $A^b_a = \delta^b_a$. Now, since the $\partial_c \delta^b_a$ vanish as well as $\partial_a g_{bc}$ one easily obtains

$$A^\alpha_a \partial_\alpha g_{\beta\gamma} + g_{\alpha\gamma}\partial_\beta A^\alpha_a + g_{\beta\alpha}\partial_\gamma A^\alpha_a = 0 \tag{14}$$

After some calculations we obtain the formulae for the connection coefficients in our coordinate system:

$$\Gamma^a_{bc} = 0, \qquad \Gamma^j_{bc} = 0, \qquad \Gamma^i_{bj} = g^{ik}g_{ab}F^a_{jk},$$
$$\Gamma^a_{bj} = g^{ik}g_{bc}A^a_k F^c_{ij} + C^a_{cb}A^c_j,$$
$$\Gamma^a_{jk} = \frac{1}{2}(\partial_j A^a_k + \partial_k A^a_j) - \left\{ \begin{matrix} m \\ jk \end{matrix} \right\} A^a_m + g^{im}g_{bc}A^a_m A^b_k f^c_{ij} + g^{im}g_{bc}A^a_m A^b_j f^c_{ik},$$

$$\Gamma^i_{jk} = \left\{ {i \atop jk} \right\} + g^{im} g_{bc} A^b_k f^c_{jm} + g^{im} g_{bc} A^b_j f^c_{km} \tag{15}$$

where we use the notation

$$f^a_{ij} = \frac{1}{2}\left(\partial_i A^a_j - \partial_j A^a_i\right) \tag{16}$$

Not all connection coefficients are explicitly covariant with respect to the gauge group; however, if the fully covariant structure constants $C_{abc} = g_{ad} C^d_{bc}$ are antisymmetric in all three indices, we can replace f^a_{ij} by the covariant quantities F^a_{ij} in the last two of connection coefficients without affecting them. In the case when g_{ab} is the Killing metric the full antisymmetry of C_{abc} takes place by virtue of the Jacobi identities.

Taking the lagrangian, by analogy with the Kaluza-Klein theory, equal to $\sqrt{-g}.R$, by well known standard procedure we obtain the Euler-Lagrange equations in the form:

$$\left(R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R\right) \delta g^{\alpha\beta} = 0 \tag{17}$$

The variations $\delta g^{\alpha\beta}$ shall not be completely arbitrary; we assume that they do not affect the particular structure of the fibre bundle. In other words, the variations $\delta g^{\alpha\beta}$ must not affect the form of the metric (13). Within these conditions the variations of g^{ij} and of A^a_i are arbitrary.

Before computing the explicit form of (17) it is useful to indentify the parts of Euler-Lagrange equations that arise from independent variations of A^a_j and g^{ij}. Replacing $g_{\alpha\beta}$ and $g^{\alpha\beta}$ by their explicit form, we obtain two systems of equations:

$$A^a_k R_{ab} - R_{bk} = 0 \tag{18}$$

corresponding to the variations of A^a_k, and

$$R_{ab} A^a_j A^b_k - 2 R_{bk} A^b_j + R_{jk} - \frac{1}{2} g_{jk} R = 0 \tag{19}$$

corresponding to the variations of g^{jk}.

Using (18) we can reduce equation (19) to the simpler form

$$R_{jk} - \frac{1}{2} g_{jk} R - A^a_j R_{ak} = 0 \tag{20}$$

Equations (18) are interpreted as the field equations, whereas (20) can be regarded as the definition of the energy-momentum tensor for the combined Yang-Mills type and gravitational fields.

After performing some computations we obtain the following expressions for the Ricci tensor:

$$\mathrm{R}_{ac} = g^{im}g^{kn}g_{ab}g_{cd}\mathrm{F}^{b}_{mn}\mathrm{F}^{d}_{ik} \tag{21}$$

$$\mathrm{R}_{bj} = g^{im}g_{ab}\nabla_{i}\mathrm{F}^{a}_{jm} - g^{il}g^{km}g_{ab}g_{cd}\mathrm{A}^{c}_{j}\mathrm{F}^{a}_{kl}\mathrm{F}^{d}_{im} - 2g^{ik}g_{bc}\mathrm{C}^{c}_{ad}\mathrm{A}^{a}_{k}\mathrm{F}^{d}_{ij} \tag{22}$$

$$\begin{aligned}\mathrm{R}_{jk} = {} & \mathrm{K}_{jk} + g^{im}g_{ab}\mathrm{A}^{a}_{j}\nabla_{i}\mathrm{F}^{b}_{km} + g^{im}g_{ab}\mathrm{A}^{a}_{k}\nabla_{i}\mathrm{F}^{b}_{jm} + g^{im}g_{ab}\mathrm{F}^{a}_{ij}\mathrm{F}^{b}_{km}\\ & + g^{im}g_{ab}\mathrm{F}^{a}_{ik}\mathrm{F}^{b}_{jm} - 2g^{im}g_{ab}\mathrm{C}^{a}_{cd}\mathrm{A}^{c}_{i}\mathrm{A}^{d}_{k}\mathrm{F}^{b}_{jm} - 2g^{im}g_{ab}\mathrm{C}^{a}_{cd}\mathrm{A}^{c}_{i}\mathrm{A}^{d}_{j}\mathrm{F}^{b}_{km}\\ & + \mathrm{C}^{d}_{ac}\mathrm{C}^{c}_{db}\mathrm{A}^{a}_{j}\mathrm{A}^{b}_{k} - g^{im}g^{ln}g_{ab}g_{cd}\mathrm{A}^{a}_{j}\mathrm{A}^{c}_{k}\mathrm{F}^{b}_{lm}\mathrm{F}^{d}_{in}\\ & + g^{im}g^{ln}g_{ab}g_{cd}\mathrm{A}^{a}_{j}\mathrm{A}^{c}_{m}\mathrm{F}^{b}_{kn}\mathrm{F}^{d}_{il} + g^{im}g^{ln}g_{ab}g_{cd}\mathrm{A}^{a}_{k}\mathrm{A}^{c}_{m}\mathrm{F}^{b}_{jn}\mathrm{F}^{d}_{il}.\end{aligned} \tag{23}$$

Here ∇_i means covariant differentiation with respect to the Christoffel symbols of the base and K_{jk} the Ricci tensor of the base. Putting these expressions into (18) we obtain

$$g^{ik}g_{ab}\nabla_{i}\mathrm{F}^{a}_{jk} = 2g^{ik}g_{ab}\mathrm{C}^{a}_{cd}\mathrm{A}^{c}_{k}\mathrm{F}^{d}_{ij}$$

or, by virtue of the non-degeneracy of the group metric tensor

$$g^{ik}\nabla_{i}\mathrm{F}^{a}_{jk} = 2g^{ik}\mathrm{C}^{a}_{bc}\mathrm{A}^{b}_{k}\mathrm{F}^{c}_{ij} \tag{25}$$

For flat space-time these equations are identical with the well known Yang-Mills equations for a non-abelian gauge field. Equation (25) generalizes them to the case of curved space-time, and also provides the form of the interaction between the gravitational and the generalized Yang-Mills field.

To obtain an explicit form of (20) and calculate the energy-momentum tensor we must compute also the Riemannian scalar R. The result is

$$\mathrm{R} = \mathrm{K} + g^{ik}g^{im}g_{ab}\mathrm{F}^{a}_{ik}\mathrm{F}^{b}_{jm} + g^{jk}\mathrm{C}^{c}_{ad}\mathrm{C}^{d}_{cb}\mathrm{A}^{a}_{j}\mathrm{A}^{b}_{k} \tag{26}$$

where K means the Riemann scalar of the base. There is always a lack of invariance due to the last term in (26); nevertheless we can see that this term vanishes by the following considerations: Since

$$\mathrm{C}^{c}_{ad}\mathrm{A}^{a}_{j} = \partial_{d}\mathrm{A}^{c}_{j} \tag{27}$$

we have

$$\begin{aligned}- g^{jk}\mathrm{C}^{c}_{ad}\mathrm{C}^{d}_{cb}\mathrm{A}^{a}_{j}\mathrm{A}^{b}_{k} &= g^{jk}(\partial_{d}\mathrm{A}^{c}_{j})(\partial_{c}\mathrm{A}^{d}_{k})\\ &= g^{ik}\partial_{d}(\mathrm{A}^{c}_{j}\partial_{c}\mathrm{A}^{d}_{k}) - g^{ik}\mathrm{A}^{c}_{j}\partial^{2}_{cd}\mathrm{A}^{d}_{k} = g^{ik}\partial_{d}(\mathrm{A}^{c}_{j}\partial_{c}\mathrm{A}^{d}_{k})\end{aligned} \tag{27}$$

because, with our assumptions concerning the full antisymmetry of C_{abc}, we have $C^d_{bd} = 0$, and then $\partial_d A^d_j = 0$.

Now

$$g^{jk}\partial_d(A^c_j\partial_c A^d_k) = \partial_d(g^{jk}A^c_j\partial_c A^d_k) = \partial_d(g^{jk}C^d_{ac}A^c_jA^a_k) = 0 \qquad (28)$$

because

$$g^{jk}C^d_{ac}A^c_jA^a_k = 0$$

The final expression for the Lagrangian is:

$$L = \sqrt{-g}R = \sqrt{-g}K + \sqrt{-g}g^{ik}g^{im}g_{ab}F^a_{ik}F^b_{jm} \qquad (29)$$

From the eq. (20) we can easily evaluate the energy momentum tensor by using the relation

$$\frac{8\pi\kappa}{c^4}T_{rs} = K_{rs} - \frac{1}{2}g_{rs}K \qquad (30)$$

where K_{rs} is the Ricci tensor of the base and κ is the gravitational constant.

From (20) we obtain:

$$\frac{8\pi\kappa}{c^4}T_{rs} = 2g^{im}g_{ab}F^a_{ri}F^b_{sm} + \frac{1}{2}g_{rs}g^{jk}g^{im}g_{ab}F^a_{ik}F^b_{jm} \qquad (31)$$

This expression is in agreement with the well known case of electromagnetism, where F^a_{jk} reduces to f_{jk} and g_{ab} to 1. The covariant divergence of T_{rs} vanishes from its definition or, if computed from the right-hand side of (31) by virtue of the field equations (25). Having at our disposal the equations (25) and (31) we can solve — at least theoretically — the problem of the interaction of a field of the Yang-Mills type with the gravitational field. Practically this can be done only with very simplifying assumptions, *e. g.* some special symmetries of the space-time and the field, or by means of perturbation calculus.

3. THE MOTION OF PARTICLES IN A FIELD OF THE YANG-MILLS TYPE AND THE GRAVITATIONAL FIELD

In analogy with the Jordan-Thirry variant of the Kaluza-Klein theory of gravitation and electromagnetism we assume, that the trajectory of a point particle is a geodesic in our M + 4 dimensional bundle manifold. Then, inserting into the equation of a geodesic the expressions for the connection coefficients (15) we obtain the equations of motion in the form:

$$\frac{d^2x^i}{ds^2} + \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} \frac{dx^j}{ds}\frac{dx^k}{ds} + 2g^{im}g_{ab}A^a_kF^b_{jm}\frac{dx^j}{ds}\frac{dx^k}{ds} + 2g^{im}g_{ab}F^b_{jm}\frac{dx^j}{ds}\frac{dx^a}{ds} = 0 \qquad (32)$$

In our coordinates the components A^a_b of the bundle connection are equal to δ^a_b, hence we can transform (32) to the form

$$\frac{d^2x^i}{ds^2} + \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} \frac{dx^j}{ds}\frac{dx^k}{ds} + 2g^{im}g_{ab}F^b_{jm}\left(A^a_\alpha \frac{dx^\alpha}{ds}\right)\frac{dx^j}{ds} = 0 \qquad (33)$$

The scalar product $A^a_\alpha \frac{dx^\alpha}{ds}$ is constant since in any metric space the geodesics are always at the constant angle to the Killing field. Introducing the notion of the generalized charge Q^a we obtain

$$\frac{d^2x^i}{ds^2} + \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} \frac{dx^j}{ds}\frac{dx^k}{ds} + g^{im}g_{ab}Q^aF^b_{jm}\frac{dx^j}{ds} = 0 \qquad (34)$$

where $Q^a = 2A^a_\alpha \frac{dx^\alpha}{ds}$. It is worthful to remark, that Q^a really does correspond to the quantity analogous to the charge divided by the mass of the particle.

The equations (34) are not complete, during the motion the charge-vector rotates in the group space. This can be seen by differentiating its definition; then the way of rotating depends explicitly on the external field.

In the case of one-dimensional abelian gauge group with g_{ab} reduced to 1 we obtain the classical Lorentz equation of an electron moving in an external field. In the non-abelian case there is an interesting new aspect to this equation of motion. Whereas in the case of the electromagnetism the necessary and sufficient conditions for the particle to move along a geodesic in space-time was either the vanishing of the charge, or the vanishing of the field, here we obtain yet another possibility: the vanishing of the group scalar product of the « charge-vector » and the field: $g_{ab}Q^aF^b_{ij} = 0$, in spite of non vanishing of both Q^a and F^b_{ij}. It is difficult to see, however, if this circumstance has any physical meaning at all.

ACKNOWLEDGEMENT

The author is greatly indebted to Professor A. Trautman for pointing out the problem, many useful discussions and criticism, and his kindly support during the preparation of this work.

REFERENCES

[*1*] Th. KALUZA, Zum Unitätsproblem der Physik. *Berl. Berichte*, 1921, p. 966.

[*2*] A. EINSTEIN, Sitz. Preuss. Akad. Wiss., phys. math., Kl 23-25, 26-30, 1927. Zu Kaluzas Theorie des Zusammenhängs von Gravitation und Elektrizität.

[*3*] C. N. YANG and R. MILLS, Conservation of Isotopic Spin and Isotopic Gauge-invariance. *Phys. Rev.*, t. **6**, 1954, p. 191.

[*4*] A. LICHNEROWICZ, *Théories relativistes de la Gravitation et de l'Électromagnétisme.* Masson et Cie, éd., Paris, 1955.

[*5*] R. UTIYAMA, Invariant theory of interactions. *Phys. Rev.*, t. **101**, 1956, p. 1597.

[*6*] A. TRAUTMAN, Lectures on Fibre Bundles, King's College, London, 1967 (mimographed).

[*7*] W. TULCZYJEW, private communication.

[*8*] H. LOOS, The range of Gauge Fields. *Nucl. Phys.*, t. **72**, 1965, p. 677.

(Manuscrit reçu le 5 juin 1968).

Non-Abelian gauge fields as Nambu-Goldstone fields*

Y. M. Cho and Peter G. O. Freund
The Enrico Fermi Institute and the Department of Physics, The University of Chicago, Chicago, Illinois 60637

(Received 31 March 1975)

Gauge invariance is obtained as a *consequence* of the spontaneous breaking of a larger symmetry. The Yang-Mills gauge fields are the corresponding Nambu-Goldstone fields. This involves the study of field theories in a space with more than four dimensions. In a "cylindrical" sector these field theories reduce to new non-Abelian generalizations of Jordan-Brans-Dicke theory. The nonobservability of the excess dimensions (while a difficulty for theories in which these dimensions are bosonic) should cause no problems if the higher dimensions are fermionic. In the latter case, field theories in which all basic fields are at the same time Nambu-Goldstone and Yang-Mills fields become possible. The relation of this work to recent work on gravitation theory by Yang is explored.

I. INTRODUCTION

Gauge fields and Nambu-Goldstone fields are the basic ingredients of unified quantum field theories of the electromagnetic, weak, and possibly even strong and gravitational interactions. With the advent of supersymmetries[1] that shuffle bosons and fermions, even Fermi fields may belong to the classes of gauge and Nambu-Goldstone fields, originally both massless (all masses in the theory are expected to originate through spontaneous symmetry breaking). One is then tempted to conjecture that these two basic categories of fields are equivalent in some sense. Maybe *every gauge field is also a Nambu-Goldstone field and/or vice versa. A priori*, the later half of this statement seems to be flawed, as for instance, the pion fields of the σ model are Nambu-Goldstone fields while they certainly are not gauge fields. Nevertheless the first half of the statement has some remarkable illustrations:

(i) The gauge field of the Abelian group U(1) or O(2) can be viewed as the Nambu-Goldstone field corresponding to the spontaneous breaking of Lorentz invariance.[2]

(ii) The gauge field of the Lorentz group (the graviton) can be viewed also as the Nambu-Goldstone field for the spontaneous breaking of general covariance.[3-5]

We shall set up a field theory within which the Yang-Mills-type gauge fields for a (non-Abelian) internal-symmetry group become Nambu-Goldstone fields for the spontaneous breaking of general covariance in a higher-dimensional space. When restricted to a 4-dimensional world this theory turns into a non-Abelian counterpart of a field theory constructed in the Abelian case by Kaluza and by Jordan.[6-8] This theory is interesting in its own right, and we work it out in detail in Secs. IV and V.

We also compare in Sec. VI our approach with Yang's recent work[9] on gravitation. Yang's theory may be more "natural" if one views gravitation as a geometric copy of Yang-Mills theory; however, if we insist on all massless fields being Nambu-Goldstone fields then Einstein's theory obtains and not Yang's.

Finally, we discuss the possibility that the higher dimensions are fermionic (in the manner of the Salam-Strathdee-Volkov-Akulov superspace[1]). In theories of this type scalar and spin-$\frac{1}{2}$ fields can appear as gauge fields, and the possibility arises that even the second half of our earlier statement holds and the concepts of gauge and Nambu-Goldstone fields become completely equivalent.

II. GRAVITATION AS A NAMBU-GOLDSTONE FIELD: A BRIEF REVIEW

Our argument that Yang-Mills fields are Nambu-Goldstone fields will exhibit a close similarity to the work of Isham, Salam, and Strathdee[3] and of Ogievetsky and Borisov[4,5] in which the Nambu-Goldstone nature of the gravitational field is established. We therefore briefly review here the major steps in their reasoning.

The best starting point is an observation due to Ogievetsky.[4] It states that any field theory in $4+N$ dimensions invariant under the general affine (i.e., inhomogeneous linear: linear + translations) group in $4+N$ dimensions $\mathrm{IGL}(4+N,R)$ and under the standard $(4+N)$-dimensional nonlinear realization of the $(4+N)$-dimensional conformal group $\mathrm{O}(4+N,2)$ is $4+N$ generally covariant. This theorem is proved by noting that the generators of $\mathrm{GL}(4+N,R)$ and of the $(4+N)$-dimensional conformal group do not close on a finite algebra, on account of the nonlinear coordinate dependence of the conformal boosts. Rather, they close on the infinite-dimensional algebra of the general covariance group in $4+N$ dimensions.

To identify gravitation with a Nambu-Goldstone boson one therefore starts from the case $N=0$ and considers the simultaneous nonlinear realization of the groups IGL($4,R$) and O(4, 2) which becomes linear for the Poincaré subgroup. Breaking O(4, 2) to the Poincaré subgroup one picks up $15-10=5$ Nambu-Goldstone fields: a scalar dilaton χ and a vector field C_m for the conformal boosts. Without loss of consistency one can, however, identify C_m with the gradient of the dilaton field, so that one picks up essentially a single scalar field this way. When spontaneously breaking IGL($4,R$) down to its Poincaré subgroup one picks up further $20-10=10$ fields which are essentially the vierbein fields in a gauge in which the vierbein matrix is symmetric. Group theory then guides the construction of what ultimately turns out to be Einstein's Lagrangian.

Alternatively one can start directly from the group of general coordinate transformations in 4-space [once one knows that this is equivalent with the GL($4,R$) – O(4, 2) procedure by Ogievetsky's theorem] and represent in a unique way any element $L^{\bar m}_m$ of GL($4,R$) in the form (Latin indices run from 1 to 4 and a bar on an index indicates that the index is "active" under Lorentz transformation)

$$L^{\bar m}_m = A^{\bar m}_{\bar a} S^{\bar a}_m . \qquad (2.1)$$

Here $L^{\bar m}_m$ is any regular 4×4 matrix, S (which plays the role of vierbein) is the exponential of a symmetric matrix, and A is a Lorentz transformation. Now consider a general (invertible) coordinate transformation

$$x'^m = x'^m(x^n) . \qquad (2.2)$$

At every point P (of old coordinates x^m and new coordinates x'^m) we associate to the transformation (2.2) the GL($4,R$) element T defined by the matrix

$$T^n_m = \frac{\partial x^n}{\partial x'^m} . \qquad (2.3)$$

To obtain the transformation law of the vierbein fields S we perform the linear transformation T on L, then decompose anew in the manner (2.1), and find the new vierbein S'. In detail,

$$L'^{\bar m}_m = L^{\bar m}_n T^n_m = A'^{\bar m}_{\bar p} S'^{\bar p}_m , \qquad (2.4)$$

so that using (2.1) we have

$$S'^{\bar m}_m = A'^{-1\bar m}_{\bar p} A^{\bar p}_{\bar a} S^{\bar a}_n T^n_m , \qquad (2.5)$$

with T^n_m given by Eq. (2.3). We stress that all fields in Eqs. (2.4) and (2.5) are evaluated at the same physical point having old coordinates x and new coordinates x', but to simplify the notation we have not explicitly written out the argument (x or x') in the various fields. Equation (2.5) establishes the transformation law of the S fields and these indeed play the role of vierbein. One then uses this vierbein to construct a metric tensor and then establishes the integral of the corresponding curvature density, i.e., the Einstein-Hilbert action

$$I_E = -\frac{1}{16\pi G}\int d^4x\sqrt{-g}\,R$$

to be an acceptable generally invariant action for the Nambu-Goldstone fields.

For our purposes, the important lesson is that the Nambu-Goldstone fields play the role of vierbein in a curved manifold and that the standard nonlinear-realization argument suffices to establish their transformation law under general coordinate transformations.

III. GAUGE FIELDS AS NAMBU-GOLDSTONE FIELDS

We want to construct a theory which exhibits manifest local gauge invariance under some N-dimensional internal symmetry G. Yet we do not want to achieve this by deliberately introducing Yang-Mills gauge fields. Rather we want these massless vector fields to emerge automatically as Nambu-Goldstone fields corresponding to the spontaneous breaking of a larger symmetry $\bar G \supset G$. Moreover, we want gauge invariance under local G transformations not to be a separate input but rather a consequence of the larger symmetry.

In implementing this program the first step must obviously be the choice of the larger group $\bar G$. To this end we first embed the N-dimensional compact group G into O(N), which is always possible. We then embed O(N) into IGL($4+N,R$) on one hand and the conformal group O($4+N$, 2) on the other hand. At the same time we extend 4-dimensional spacetime to a ($4+N$)-dimensional space with one timelike and $3+N$ spacelike dimensions. By Ogievetsky's theorem a field theory in ($4+N$)-dimensional space invariant under IGL($4+N,R$) and O($4+N$, 2) is invariant under the general coordinate transformations in the ($4+N$)-dimensional space. Now, let us follow the 4-dimensional procedure and look for a nonlinear realization of($4+N$)-general covariance which becomes linear for the ($4+N$)-dimensional Poincaré group P_{4+N}. The dimensions of IGL($4+N,R$), O($4+N$, 2), and P_{4+N} are respectively $(4+N)(5+N)$, $(5+N)(6+N)/2$, and $(4+N)(5+N)/2$. From the breaking of the conformal group we then pick up $(5+N)(6+N)/2-(4+N)(5+N)/2=N+5$ Nambu-Goldstone fields: one scalar and one ($4+N$)-vector. The latter is again the ($4+N$) gradient of

SPACE-TIME GEOMETRY AND SYMMETRY BREAKING

J.F. LUCIANI
Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, Paris, France *

Received 8 September 1977

We look for solutions of the Einstein-Yang-Mills equations in a $4 + D$ dimensional space-time. We find solutions where the first 4 dimensions are a flat Minkowskian space-time, while the D others are a compact, space-like manifold of small size. Such solutions can be obtained for an arbitrary compact gauge group K and are invariant under a subgroup G of K related to the space-time geometry. This shows that $4 + D$ dimensional gravity can give a mechanism for the super-strong symmetry breaking needed in grand unified field theories without introducing Higgs scalars.

1. Introduction

The idea of using an internal space to generate symmetries dates at least from 1921, when Kaluza [1] gave a geometrical treatment of the Maxwell-Einstein theory, in the framework of a five-dimensional space. More recently, it has been possible to generalize the idea to an arbitrary gauge group [2]. But this requires the introduction of many extra dimensions, and of a specific structure for space-time, namely a fibre bundle. Thus, the extra dimensions have lost their physical sense as real space-time dimensions.

In our work, we investigate a quite different problem: we start with a theory in which the equations of motion in the $4 + D$ dimensions are symmetric, namely, an Einstein-Yang-Mills theory, and we search for solutions where one has a D-dimensional internal space, compact, and whose dimensions are small. We do this for two purposes: first, to give a physical meaning to theories containing gravitation and gauge fields in a $4 + D$ dimensional space, like the spinor dual model [3] for which $D = 6$, or for a supergravity theory in 10 dimensions, whose existence is indicated by the spinor dual model [4].

Secondly, unified gauge theories [5] of quarks and leptons need a spontaneous symmetry breaking giving rise to huge masses for leptoquarks-type bosons. From the lifetime of the proton, and from asymptotic freedom considerations, one can evaluate that the typical order of magnitude needed in a realistic model may be as large as Planck's mass (10^{19} GeV). Such a huge breaking is very hard to achieve by the means of Higgs scalars, as shown recently by Gildener [6]. So, it was speculated by Georgi, Quinn and Weinberg [7] that gravity may have something to do with this huge symmetry breaking. In a recent paper it was shown [8] that in the framework of a unified gauge theory in a 6-dimensional space-time, this symmetry breaking could indeed arise out of the mechanism of spontaneous compactification. We generalize this idea to the general $4 + D$ dimensional case. The first solutions of the $4 + D$ dimensional Yang-Mills-Einstein equations where spontaneous compactification occurred were obtained by Cremmer and Scherk [9], in the case of an $SO(D + 1)$ gauge group, the internal space being an S_D sphere. We generalize their solutions for a large class of internal spaces, realized as homogeneous spaces on a group G, such that any simple compact Lie group admits at least one such space. The "size" of the internal space is given by one parameter, of the order of Planck's length (10^{-33} cm). Thus, by expanding any field on the internal space in series of harmonics, we shall get a four-dimensional spectrum of masses classified by G, with only zero and superheavy masses. Only the zero-mass states are observable at usual energies.

---

* Laboratoire propre du CNRS, Associé à l'Ecole Normale Supérieure et à l'Université de Paris-Sud. Postal address: 24, rue Lhomond, 75231 Paris Cédex 05, France.

We shall show further that the group G not only classifies the states, but is also a local gauge group, i.e. that when one expands the fields around our classical solutions, taken as a possible "vacuum" of the system, 4-dimensional massless gauge bosons associated with the group G appear in the spectrum.

Sect. 3 of our work can be considered as a mathematical extension of the work done on the fibre bundle approach [2] to the unification of gravity and gauge fields, where instead of working with the group space, we work with homogeneous spaces. In our case, the fibre bundle structure of space-time is not imposed but arises as a classical solution of the field equations.

Our models may look less geometrical than the fibre bundle approach since the starting point (Einstein-Yang-Mills theory) is not a unified geometrical theory. On the other hand the situation has changed recently with the advent of extended supergravity theories [10] where the Einstein-Yang-Mills theory appears as the bosonic sector of a unified field theory involving fermions (spin $\frac{3}{2}$ and 1). Also in the context of the spinor dual model [4], the Einstein-Yang-Mills fields appear as a subsector of a unified theory.

Finally, the non-renormalizability of $4 + D$ dimensional theories is not a real drawback in our case, since we merely want to obtain an understanding of how the extra dimensions in a dual spinor model can be made unobservable and used to generate a huge symmetry breaking. In such a realistic model, the Einstein-Yang-Mills theory is automatically short-distance modified and renormalizability is achieved.

2. Compactification of an homogeneous space

2.1. Generalities

We start with a theory of gravitation, in interaction with gauge fields (gauge group K), with a cosmological term, in the $(4 + D)$-dimensional space E:

$$\mathcal{L} = -\frac{1}{16\pi G}\sqrt{-g}\,R - \frac{1}{4e^2}\sqrt{-g}\,G^\alpha_{\hat\mu\hat\nu}\,G^{\hat\mu\hat\nu\alpha} - \sqrt{-g}\,V_0\,,$$

$$G^\alpha_{\hat\mu\hat\nu} = \partial_{\hat\mu}B^\alpha_{\hat\nu} - \partial_{\hat\nu}B^\alpha_{\hat\mu} + f^{\alpha\beta\gamma}B^\beta_{\hat\mu}B^\gamma_{\hat\nu}\,.$$

The $\hat\mu$ indices will always refer to the whole space E (whose metric has a signature $-1, +1, ..., +1$); the structure constants are normalized to:

$$f^{\alpha\beta\gamma}f^{\alpha'\beta\gamma} = \delta^{\alpha\alpha'}\,.$$

They are antisymmetric on the three indices. (The group K is simple and compact). We assume that, in a large enough domain, E can be split as:

$$\mathrm{E} = \mathrm{M} \times \mathrm{S}\,, \qquad \begin{array}{ll} \text{M:} & \text{Minkowski space,} \\ \text{S:} & \text{compact, space-like space.} \end{array}$$

Relatively to this decomposition, we shall take coordinates (x^μ, θ^i), μ indices for M, i for S.

The equations of motion are:

$$R_{\hat\mu\hat\nu} - \tfrac{1}{2}Rg_{\hat\mu\hat\nu} = -8\pi g\, T_{\hat\mu\hat\nu}\,, \qquad T_{\hat\mu\hat\nu} = \frac{1}{e^2}\,G^\alpha_{\hat\mu\hat\rho}\,G^{\alpha\hat\rho}_{\hat\nu} - g_{\hat\mu\hat\nu}(G^\alpha_{\hat\rho\hat\lambda}G^{\alpha\hat\rho\hat\lambda} + V_0)\,,$$

$$\hat D_{\hat\mu}G^{\hat\mu\hat\nu} = 0\,,$$

where $\hat D_{\hat\mu}$ is the Yang-Mills covariant derivative. We search for a solution where:

$$g_{\hat\mu\hat\nu} = \begin{bmatrix} g_{\mu\nu}(x^\lambda) & 0 \\ & \\ 0 & g_{ij}(x^k) \end{bmatrix}$$

and $R_{\mu\nu} = 0$. This includes the possibility for the M space to have a metric corresponding to freely travelling gravitational waves. It is natural to make the ansatz:

$$B^\alpha_\mu = 0\,,$$

$$B^\alpha_i = B^\alpha_i(\theta^j)\,.$$

Thus the equations become:

$$-\tfrac{1}{2}R^i_i g_{\mu\nu} = 8\pi G\, g_{\mu\nu}\left(\frac{1}{4e^2}\,G^\alpha_{ij}\,G^{\alpha ij} + V_0\right),$$

$$R_{ij} - \tfrac{1}{2}R^k_k g_{ij} = -8\pi G\left[\frac{1}{e^2} G^\alpha_{ik}\, G^{\alpha k}_j - g_{ij}\left(\frac{1}{4e^2} G^\alpha_{kl}\, G^{\alpha kl} + V_0\right)\right] ,$$

$$\hat{D}_i G^{\alpha ij} = 0 .$$

So we are left with:

$$R_{ij} = -\frac{8\pi G}{e^2} G^\alpha_{ik}\, G^{\alpha k}_j , \tag{1}$$

$$V_0 = -\frac{1}{4 \times 8\pi G} R^i_i , \tag{2}$$

$$\hat{D}_i G^{\alpha ij} = 0 . \tag{3}$$

We shall now forget for a moment the M space and seach for solutions of (1) and (3) on the S space.

Note that, for a solution of this type, the action vanishes, since:

$$-\tfrac{1}{2}R^i_i = 8\pi G \frac{1}{4e^2} G^\alpha_{ij}\, G^{\alpha ij} + V_0 \quad .$$

So, these solutions can be called "instantons" in a generalized way. Since the metric of S is assumed to be space-like (for the positivity of the spectrum of masses), R^i_i is constant and negative, and V_0 is positive as in the super-symmetric theories where a "mass" term for the spin-$\frac{3}{2}$ fields is introduced [10]. For such theories, V_0 is constrained, and has a form similar to the one we shall find.

2.2. Form of the solution

In order to find a solution for eqs. (1) and (3), subject to the constraint (2), we shall suppose that S is a homogeneous space on a compact simple group G, that is, we can define an action of G on S such that any two points of S can be connected by an element of G. Furthermore, we shall assume in sects. 2 and 3 that the gauge group K is G, and return to the general case in sect. 4. Later, we shall investigate the mathematical structure of these spaces. Now, we only need a few properties stated as follows:

(a) The action of G can be represented by differential operators L^α, with: $[L^\alpha, L^\beta] = f^{\alpha\beta\gamma}L^\gamma$ (again, $f^{\alpha\beta\gamma}$ are normalized).

Writing $L^\alpha = A^{i\alpha}\partial_i$, we get:

$$A^{i\alpha}\partial_i A^{j\beta} - A^{i\beta}\partial_i A^{j\alpha} = f^{\alpha\beta\gamma}A^{j\gamma} . \tag{4}$$

(b) On S, we can build an invariant metric (with upper indices):

$$g^{ij}_0(\theta) = A^{i\alpha}A^{j\alpha} .$$

The metric is G-invariant, that is, if $\theta^i \to \theta'^i = R(\theta^j)$ is a general coordinate transformation induced by an element of G:

$$g_0^{ij}(\theta'^k) = \frac{\partial\theta'^i}{\partial\theta^k}\frac{\partial\theta'^j}{\partial\theta^l}\, g_0^{kl}(\theta^n) \tag{5}$$

(to prove it, we need only (4)).

(c) Then the $A^{i\alpha}$ are Killing vectors:

$$\nabla_i A_j^\alpha + \nabla_j A_i^\alpha = 0\,. \tag{6}$$

(d) We introduce

$$h^{\alpha\beta} = A^{i\alpha} A_i^\beta\,.$$

Then,

$$h^{\alpha\beta} h^{\beta\gamma} = h^{\alpha\gamma}$$

and

$$A^{i\alpha}\partial_i h^{\beta\gamma} = f^{\alpha\beta\delta} h^{\delta\gamma} + f^{\alpha\gamma\delta} h^{\beta\delta} \qquad \text{(results from (4) and (6))}.$$

Note that, in the particular case S = G, $h^{\alpha\beta} = \delta^{\alpha\beta}$.

Now, we make the ansatz:

$$B^{i\alpha} = \lambda A^{i\alpha}\,,$$

$$g^{ij} = \frac{1}{\mu} g_0^{ij} = \frac{1}{\mu\lambda^2} B^{i\alpha} B^{j\alpha}\,.$$

μ is related to the parameter R_0 having the dimension of a length introduced in ref. [9], and characterizes the size of the internal space by: $\mu = R_0^2/2(D-1)$. It will be determined by the equations of motion. λ is essentially μ^{-1} up to a dimensionless constant to be fixed. Inspecting eq. (1), we see that the dimensionless quantity $8\pi G/e^2$ appears everywhere, so that μ will always be given by $\mu \sim 8\pi G/e^2$ (up to a numerical constant).

We begin with the calculation of $\nabla_i B_j^\alpha$. From

$$B^{i\alpha}\nabla_i B^{j\beta} - B^{i\beta}\nabla_i B^{j\alpha} = \lambda f^{\alpha\beta\gamma} B^{j\gamma}\,,$$

we get (from eq. (6))

$$\nabla_i B_j^\epsilon(\delta^{\alpha\epsilon} + h^{\alpha\epsilon}) = \frac{1}{\lambda\mu} f^{\beta\alpha\gamma} B_j^\gamma B_i^\beta\,.$$

The inverse of $\delta^{\alpha\beta} + h^{\alpha\beta}$ is $\delta^{\alpha\beta} - \frac{1}{2}h^{\alpha\beta}$ (because $h^{\alpha\beta}h^{\beta\gamma} = h^{\alpha\gamma}$). Then

$$\nabla_i B_j^\alpha = \frac{1}{\lambda\mu} f^{\beta\epsilon\gamma} B_i^\beta B_j^\gamma (\delta^{\alpha\epsilon} - \tfrac{1}{2}h^{\alpha\epsilon})\,.$$

Now, we can find the Ricci tensor, by

$$\nabla^i \nabla_i B_j^\alpha = -\nabla^i \nabla_j B_i^\alpha = -\nabla_j \nabla^i B_i^\alpha + R_j^k B_k^\alpha ,$$

$$\nabla^i \nabla_i B_j^\alpha = -R_j^k B_k^\alpha \quad , \qquad \text{from (6)} .$$

A direct calculation, using the preceding formulae, gives:

$$R_{ij} = -\left(\frac{3}{4\mu} g_{ij} - \frac{1}{2\lambda^2 \mu^2} B_i^\alpha B_j^\beta K^{\alpha\beta}\right) .$$

The tensor $K^{\alpha\beta}$ will have a great importance. It is given by $K^{\alpha\beta} = f^{\alpha\gamma\delta'} h^{\delta\delta'}$. Now, we can easily compute:

$$G_{ij}^\alpha = 2\nabla_i B_j^\alpha + f^{\alpha\beta\gamma} B_i^\beta B_j^\gamma \qquad \text{and} \qquad \hat{D}_i G^{ij\alpha} ,$$

by computing the four terms appearing in it. For example:

$$B^{i\beta} \nabla_i B^{j\gamma} f^{\alpha\beta\gamma} = \tfrac{1}{2}\lambda\, f^{\alpha\beta\gamma} f^{\beta\gamma\delta} B^{j\delta} = \tfrac{1}{2}\lambda\, B^{j\alpha} .$$

In the same way, in the calculation of $G_{ik}^\alpha G_j^{\alpha k}$, we use:

$$(\nabla_i B_k^\alpha)(\nabla_j B^{\alpha k}) = -\lambda^2 \mu\, R_{ij}$$

(which also results from the preceding formulae). Thus, we get:

$$\hat{D}_i G^{ij\alpha} = (1 - \lambda\mu)\left(\frac{1 + \lambda\mu}{\mu} K^{\alpha\beta} B^{j\beta} - \frac{3}{2\mu} B^{j\alpha}\right) ,$$

$$G_{ik}^\alpha G_j^{\alpha k} = g_{ij}(3\lambda^2 - \lambda^3 \mu) + B_i^\alpha B_j^\beta K^{\alpha\beta}\left(\lambda^2 \mu - \frac{2}{\mu}\right) .$$

We can now solve eqs. (1) and (3):

(a) If $K^{\alpha\beta} B_i^\beta \neq C\, B_i^\alpha$, the Yang-Mills equation admits the solution $1 - \lambda\mu = 0$ only. But we cannot solve eq. (1). Our ansatz cannot be a solution.

(b) Now the interesting case is when

$$K^{\alpha\beta} B_i^\beta = C\, B_i^\alpha \tag{7}$$

(C is a constant).

This is constraint on the geometrical structure of the space S, and this will be solved in subsect. 2.3. Assuming it to be realized, we find that the Yang-Mills equation admits two solutions:

(α) $\lambda\mu = 1$. Then,

$$G_{ik}^\alpha G_j^{\alpha k} = \lambda^2 g_{ij}(1 - C) ,$$

$$-R_{ij} = \frac{1}{\mu} g_{ij}(\tfrac{3}{4} - \tfrac{1}{2}C) ,$$

and

$$\mu = \frac{8\pi G}{e^2}\,\frac{1-C}{\frac{3}{4}-\frac{1}{2}C}\,, \qquad V_0 = \frac{e^2}{4\times(8\pi G)^2}\,\frac{D(\frac{3}{4}-\frac{1}{2}C)^2}{1-C}\,.$$

This solution is possible if $C \neq 1$ (we always have $C \leqslant 1$). Note that V_0 is constrained.

(β) $\lambda\mu = \dfrac{3}{2C} - 1$. Then,

$$G^{\alpha}_{ik}G^{\alpha k}_{j} = \lambda^2 g_{ij}(\tfrac{3}{4} - \tfrac{1}{2}C)(C - \tfrac{1}{2})\,\frac{2}{C}\,,$$

$$-R_{ij} = \frac{1}{\mu}\,g_{ij}(\tfrac{3}{4} - \tfrac{1}{2}C)\,,$$

$$\mu = \frac{8\pi G}{e^2}\,2(C - \tfrac{1}{2})\left(\frac{3}{2C} - 1\right)^2, \qquad V_0 = \frac{e^2}{4\times(8\pi G)^2}\,\frac{DC}{4(C-\frac{1}{2})(\frac{3}{2}C-1)}\,.$$

Thus, we need $\mu > 0$, so $C > \frac{1}{2}$, otherwise the S space would be time-like, and negative (mass)^2 would appear in the spectrum of states. We shall see in sect. 3 that, though G and e are not the constants observed in the 4 dimensions, the ratio $8\pi G/e^2$ is of the same order as $8\pi G^{(4)}/e^{(4)2}$. Thus, we get $R_0 \sim 10^{-33}$ cm (for $e^{(4)} \sim 1$).

2.3. *Homogeneous spaces*

In order to see if the crucial condition (7) can be fulfilled, we must investigate the mathematical structure of homogeneous spaces [11]. Let us denote by $g \cdot x$ the action of an element g of G on an element x of S. For a given point x_0 in S, let us call H the subgroup of G such that:

$$\mathrm{H} \cdot x_0 = x_0\,.$$

Then S can be realized as the cosets $g\mathrm{H} = \{gh\}$ for $h \in \mathrm{H}$ (i.e. by identifying all the elements of G of the form $g \cdot h$, for a given g and arbitrary $h \in \mathrm{H}$).

Furthermore, we have an application π of G to S, such that: $\pi(g) = g \cdot \mathrm{H}$. Thus,

$$\pi(e) = x_0\,,$$

$$\pi(gh) = \pi(g) \qquad \text{for } h \in \mathrm{H}\,.$$

We recall that if π is an application of E to F, its differential $\mathrm{d}\pi$ is defined by:

$$\mathrm{d}\pi^i_k = \frac{\partial(\pi(x))^i}{\partial x^k}\,,$$

and that the Lie algebra of G can be viewed as the set of right-invariant vector fields on G. Let us denote by C^α a basis of the Lie algebra $\mathcal{G}$ of G, constituted by such fields. Since the fields C^α are right-invariant, we can use $\mathrm{d}\pi$ to transport them on S. Then we have:

$$A^{i\alpha} = (\mathrm{d}\pi)^i_k C^{k\alpha} \; .$$

The $A^{i\alpha}$ are just the fields introduced above (this is sufficient to verify eq. (4)). Let us give an explicit example: G is SU(2), S is a 2-dimensional sphere S_2. Thus $H = U(1)$. $\left(\begin{smallmatrix}\alpha & \beta \\ \gamma & \delta\end{smallmatrix}\right)$ being a matrix of SU(2), let us set:

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} .$$

Then (Z_1, Z_2) verifies: $|Z_1|^2 + |Z_2|^2 = 1$. Furthermore, the application $\left(\begin{smallmatrix}\alpha & \beta \\ \gamma & \delta\end{smallmatrix}\right) \to \left(\begin{smallmatrix}Z_1 \\ Z_2\end{smallmatrix}\right)$ is one to one. Thus, SU(2) is the sphere S_3.

We can parametrize S_3 by:

$$Z_1 = \mathrm{e}^{i\varphi_1} \cos \tfrac{1}{2}\theta' \, ,$$

$$Z_2 = \mathrm{e}^{i\varphi_2} \sin \tfrac{1}{2}\theta' \, .$$

Then, the fields $C^{i\alpha}$ generate the rotations on S_3 (the generators are $R^\alpha = C^{i\alpha}\partial_i$). One finds, in the basis of coordinates $(\theta', \varphi_1, \varphi_2)$:

$$C_z = -\tfrac{1}{2}(0, 1, -1) \, ,$$

$$C_+ = -\frac{i}{2\sqrt{2}} \mathrm{e}^{i(\varphi_1 - \varphi_2)} (1, -i \operatorname{tg} \tfrac{1}{2}\theta' , -i \operatorname{cotg} \tfrac{1}{2}\theta') \, ,$$

$$C_- = -\frac{i}{2\sqrt{2}} \mathrm{e}^{-i(\varphi_1 - \varphi_2)} (1, i \operatorname{tg} \tfrac{1}{2}\theta' , i \operatorname{cotg} \tfrac{1}{2}\theta') \, .$$

But SU(2) can also be parametrized by three Euler angles (φ, θ, ψ):

$$g(\varphi, \theta, \psi) = R_2(\varphi)\, R_x(\theta)\, R_2(\psi) \, .$$

So, the cosets $g \cdot H$, where H is the $(0, 0, \psi)$ rotations, parametrize the sphere S_2: (φ, θ) are the polar angles, and the projection is just:

$$\pi(\varphi, \theta, \psi) = (\varphi, \theta) \, .$$

The correspondence between the two sets of coordinates on SU(2) is:

$$\theta' = \theta \, ,$$

$$\varphi_1 = -\tfrac{1}{2}(\varphi + \psi) \, ,$$

$$\varphi_2 = -\tfrac{1}{2}(\psi - \varphi) - \tfrac{1}{2}\pi \, .$$

So, we can find C^α in the (θ, φ, ψ) base:

$$C_z = \tfrac{1}{2}(0, 1, 0)\,,$$

$$C_+ = \tfrac{1}{2}\sqrt{\tfrac{1}{2}}\; e^{i\varphi}(1, i\,\mathrm{tg}\,\theta, i/\sin\theta)\,,$$

$$C_- = \tfrac{1}{2}\sqrt{\tfrac{1}{2}}\; e^{-i\varphi}(-1, i\,\mathrm{tg}\,\theta, i/\sin\theta)\,,$$

(Note that C^α, being right-invariant, do not depend on ψ). We obtain (on the sphere in the (θ, ψ) basis):

$$A_x = \tfrac{1}{2}(-\sin\varphi, \cos\varphi\,\mathrm{tg}\,\theta)\,,$$

$$A_y = \tfrac{1}{2}(\cos\varphi, -\sin\varphi\,\mathrm{tg}\,\theta)\,,$$

$$A_z = \tfrac{1}{2}(0, -1)\,.$$

This is the ansatz of refs [9]. (It is exactly the angular dependence of the 't Hooft monopole [12]).

Now, we can investigate the structure of $K^{\alpha\beta}$. (Let us denote by $\mathcal{H}$ the Lie algebra of H, and write: $\mathcal{G} = \mathcal{H} \oplus \mathcal{U}$. $\mathcal{U}$ is the space orthogonal to $\mathcal{H}$ for the Killing metric, that we shall denote by $(\alpha|\beta)$. We shall also denote $\alpha \in \mathcal{H}$ or $\alpha \in \mathcal{U}$, for a direction C^α of the Lie algebra in $\mathcal{H}$ or $\mathcal{U}$). Then, since $\mathrm{H}\cdot x_0 = x_0$, we have:

$$A^{i\alpha}(x_0) = \mathrm{d}\pi^i_k C^{k\alpha}(e) = 0 \qquad \text{for } \alpha \in \mathcal{H}\,,$$

$$A^{i\alpha}(x_0)\cdot A^\beta_i(x_0) = \mathrm{S}^{\alpha\beta} \qquad \text{for } \alpha, \beta \in \mathcal{U}\,. \tag{8}$$

Then, at the point x_0:

$$K^{\alpha\beta} = f^{\alpha\gamma\delta} f^{\beta\gamma\epsilon} h^{\epsilon\delta} = \delta^{\alpha\beta} - \sum_{\delta \in \mathcal{H}} f^{\alpha\gamma\delta} f^{\beta\gamma\delta}\,.$$

Any element X in $\mathcal{H}$ acts on $\mathcal{G}$, by the relation:

$$\mathrm{ad}X \cdot Y = [X, Y]$$

(in particular, for the subspace $\mathcal{H}$ of $\mathcal{G}$ we get the adjoint representation of $\mathcal{H}$).

If we call Δ the Casimir operator: $\Sigma_{\delta\in\mathcal{H}}\, (\mathrm{ad}\, x^\delta)^2$ we have:

$$K^{\alpha\beta} = \delta^{\alpha\beta} - (\alpha|\Delta|\beta)\,.$$

So, at x_0, the condition (7) can be written (because of (8))

$$(\alpha|\Delta|\beta) = (1 - C)\,\delta^{\alpha\beta} \qquad \text{for } \alpha, \beta \text{ in } \mathcal{U}\,,$$

$$(\alpha|\Delta|\beta) = 0 \qquad \text{for } a \text{ in } \mathcal{U}\,, \qquad \beta \text{ in } \mathcal{H}\,,$$

$(\alpha|\Delta|\beta)$ arbitrary in $\mathcal{H}$.

The second relation is trivially verified, because:

$$[\mathcal{H}, \mathcal{U}] \subset \mathcal{U} \ ,$$

$$[\mathcal{H}, \mathcal{H}] \subset \mathcal{H} \ .$$

Furthermore, we can make the same analysis at any point x of S, by a rotation of the Lie algebra $\mathcal{G}$, such that:

$$A^{i\alpha} = 0 \ , \qquad \alpha \in \mathcal{H} \ ,$$

$$A^{i\alpha} A_i^{\beta} = \delta^{\alpha\beta} , \qquad \alpha, \beta \in \mathcal{U} \ ,$$

where $\mathcal{H}'$ is the Lie algebra of the group H′ corresponding to x. (We recall that H′ = gHg^{-1} for some g in G).

So a sufficient condition in order to fulfill (7) is:

$$(\alpha|\Delta|\beta) = (1 - C)\,\delta^{\alpha\beta} \qquad \text{for } \alpha, \beta \text{ in } \mathcal{U} \ .$$

We shall give two general types of spaces where it works:

(i) If $\mathcal{H} = 0$, S is a (simple compact) group. Then $\Delta = 0$, and $C = 1$. In this case, we must choose the solution

$$\lambda\mu = \tfrac{1}{2} \ , \qquad \mu = \frac{8\pi G}{4e^2} \ , \qquad V_0 = \frac{D}{4} \frac{e^2}{(\gamma\pi G)^2} \ .$$

(ii) If the Lie algebra $\mathcal{G}$ has an involutive automorphism σ:

$$\sigma[x, y] = [\sigma(X), \sigma(Y)] \ ,$$

$$\sigma^2 = 1 \ .$$

Then we get a subalgebra $\mathcal{H}$ such that $\sigma = 1$ on $\mathcal{H}$, and its orthogonal $\mathcal{U}$, on which $\sigma = -1$. Furthermore:

$$[\mathcal{H}, \mathcal{H}] \subset \mathcal{H} \ , \qquad [\mathcal{H}, \mathcal{U}] \subset \mathcal{U} \ , \qquad [\mathcal{U}, \mathcal{U}] \subset \mathcal{H} \ .$$

We can construct the homogeneous space G/H (H has the algebra $\mathcal{H}$). Then, for this space:

$$\underset{\alpha, \beta \in \mathcal{U}}{(\alpha|\Delta|\beta)} = \sum_{\substack{\gamma \in \mathcal{H} \\ \delta \in \mathcal{G}}} f^{\alpha\gamma\delta} f^{\beta\gamma\delta} = \sum_{\substack{\gamma \in \mathcal{H} \\ \delta \in \mathcal{U}}} f^{\alpha\gamma\delta} f^{\beta\gamma\delta} \qquad \text{because } [\mathcal{U}, \mathcal{H}] \subset \mathcal{U}$$

$$= \sum_{\delta \in \mathcal{U}} f^{\alpha\gamma\delta} f^{\beta\gamma\delta} \qquad \text{because } [\mathcal{U}, \mathcal{U}] \subset \mathcal{H}$$

$$= \delta^{\alpha\beta} - (\alpha|\Delta|\beta) \ .$$

So

$$(\alpha|\Delta|\beta) = \tfrac{1}{2}\delta^{\alpha\beta} \qquad \text{for } \alpha, \beta \in \mathcal{U}\ .$$

Then the condition (7) is fulfilled, with $C = \frac{1}{2}$. For these spaces we must take the solution $\lambda\mu = 1$ and we get:

$$\mu = \frac{8\pi G}{e^2}\,, \qquad V_0 = \frac{D}{8}\,\frac{e^2}{(8\pi G)^2}\ .$$

A classification of these automorphisms of simple Lie algebras has been made. (All the simple Lie algebras admit such automorphisms.) For the usual groups, we get:

$$\mathrm{G} = \mathrm{SU}(P+Q)\,, \qquad \mathrm{H} = \mathrm{SU}(P) \times \mathrm{SU}(Q) \times \mathrm{U}(1)\,, \qquad \dim \mathrm{S} = 2PQ\,,$$

$$\mathrm{G} = \mathrm{SU}(N)\,, \qquad \mathrm{H} = \mathrm{SO}(N)\,, \qquad \dim \mathrm{S} = \tfrac{1}{2}(N-1)(N+2)\,,$$

$$\mathrm{G} = \mathrm{SO}(2N)\,, \qquad \mathrm{H} = \mathrm{U}(N)\,, \qquad \dim \mathrm{S} = N(N-1)\,,$$

$$\mathrm{G} = \mathrm{SO}(P+Q)\,, \qquad \mathrm{H} = \mathrm{SO}(P) \times \mathrm{SO}(Q)\,, \qquad \dim \mathrm{S} = PQ\ .$$

The solution previously found by Cremmer and Scherk [8] corresponds precisely to $\mathrm{G} = \mathrm{SO}(N+1)$, $\quad \mathrm{H} = \mathrm{SO}(N)$.

Note the interesting feature that the two values $C = 1$ and $C = \frac{1}{2}$ are just the critical values of the Einstein equations.

We do not know if the condition (7) can be fulfilled by another type of space. Very likely, it seems that it is not the case. It is easy to find examples in which it is not satisfied. One such example is G = SU(3), H = SU(2). Note that the groups G, H and spaces S can be replaced by locally identical groups and spaces. This can change the spectrum of masses. In the same way, we can replace G by a product of simple groups. For S, we can thus obtain a product of spaces, or something more complicated (with, again, a different mass spectrum).

If condition (7) is not fulfilled, a solution might be obtained by writing:

$$B^{i\alpha} = \Lambda^{\alpha\beta} A^{i\beta}\,,$$

$$g^{ij} = M^{\alpha\beta} A^{i\alpha} A^{j\beta}\,,$$

where $\Lambda^{\alpha\beta}$ and $M^{\alpha\beta}$ are constant in each irreducible representation of $\mathcal{H}$ in $\mathcal{G}$.

3. Local gauge invariance

The solution we have written can be considered as the classical vacuum state around which the theory will be quantized. It was shown in ref. [13] that one can first reformulate the theory by expanding all fields around our classical solution, and integrating the action over the coordinates of the S space. Then one obtains a 4-dimensional field theory, with an infinite set of fields, of spins 2, 1 and 0 in this case. The squared masses of these fields are quantized and are given by N/μ, N being dimensionless quantities related to the eigenvalues of the Casimir operator of G. As μ turns out to be very small, we are mostly concerned with exhibiting the field content of the theory for $N = 0$, the fields of $(\text{mass})^2$ N/μ, with $N \neq 0$ (for instance ultragravitons) being unobservable at present energy.

As shown in ref. [9] one of the fields with $N = 0$ is the ordinary gravitational field. As we have a vacuum expectation value to the gauge fields, one may wonder whether or not some gauge fields with zero mass still subsist. We shall prove that, though we have broken the symmetry by choosing a solution of the equations of motion, there remain a local invariance and a set of gauge fields of the group G. The following is only a generalization for homogeneous spaces of the formalism of fibre bundles used to unify gravitation and gauge theories.

For the sake of simplicity, we shall suppose that S is a space of type II, and we let $8\pi G/e^2 = 1$, so that $\lambda = \mu = 1$. (We shall restore the units at the end.).

The main fact is that the solution found above is invariant under a rotation of G, followed by the same rotation in the Lie algebra of K (here, K = G) for the gauge fields B^α_i. Indeed: g_{ij} is invariant (eq. (5)).

If we make the coordinate transformation $\theta'^i = \theta^i + \Lambda^\alpha A^{i\alpha}$ we get:

$$A'^{\beta}_i(\theta'^k) = \frac{\partial\theta^j}{\partial\theta'^i} A^\beta_j(\theta^k) ,$$

$$\delta A^\beta_i(\theta'^k) = -\Lambda^\alpha(\partial_i A^{j\alpha})\, A^\beta_j - \Lambda^\alpha A^{j\alpha}\partial_j A^\beta_i$$

$$= \Lambda^\alpha(A^{j\beta}\nabla_j A^\alpha_i - A^{j\alpha}\nabla_j A^\beta_i) = -\Lambda^\alpha f^{\alpha\beta\gamma} A^\gamma_i .$$

Thus, we can write:

$$(R + T)\, A^\beta = 0 ,$$

where R is the above transformation, and

$$(TA_i)^\beta = \Lambda^\alpha f^{\alpha\beta\gamma} A^\gamma_i ,$$

In fact, the transformation is defined only for the D-dimensional space. In the whole space, a global rotation is defined only by a choice of coordinate (x^μ, θ^i). If we change (x^μ, θ^i) into $(x^\mu, R(\theta^i))$, where R is a local rotation, we get another decomposition of E, and we can find another solution relative to this new choice. Indeed, it will have the same form, but it will not be the same.

This indicates that we have a gauge theory associated with these local rotations. In order to find the gauge bosons of zero mass, consider the metric:

$$g_{\hat\mu\hat\nu} = \begin{bmatrix} g_{\mu\nu}(x^\lambda) + g_{\mu i} g_{\nu j} g^{ij} & g_{\mu j}(x, \theta) \\ g_{i\nu}(x, \theta) & g_{ij}(\theta) \end{bmatrix} ,$$

g_{ij} is the same as before $g_{i\mu} = W^{\alpha}_{\mu}(x)\, A^{\alpha}_{i}(\theta)$.

If we make a change of coordinates: $(x'^{\mu}, \theta'^{i}) = (x^{\mu}, \theta^{i} + \Lambda^{\alpha}(x)\, A^{i\alpha})$ we get:

$$g'_{i\mu}(x', \theta') = \frac{\partial x^{\hat{\mu}}}{\partial \theta'^{i}} \frac{\partial x^{\hat{\rho}}}{\partial x'^{\mu}}\, g_{\hat{\mu}\hat{\rho}}(x, \theta)$$

$$= \frac{\partial \theta^{j}}{\partial \theta'^{i}}\, g_{\mu j}(x, \theta) + \frac{\partial \theta^{j}}{\partial x'^{\mu}}\, g_{ij}(\theta) ,$$

$$\delta g_{i\mu}(x', \theta') = -\Lambda^{\alpha} [(\partial_{i} A^{j\alpha})\, A^{\beta}_{j} + A^{j\alpha} \partial_{j} A^{\beta}_{i}]\; W^{\beta}_{\mu} - \partial_{\mu} \Lambda^{\alpha} A^{\alpha}_{i} ,$$

so we can write

$$\delta g_{i\mu}(x, \theta) = A^{\alpha}_{i}(\theta)\, \delta W^{\alpha}_{\mu}(x) ,$$

with

$$A^{\alpha}_{i} \delta W^{\alpha}_{\mu}(x) = -\Lambda^{\alpha} f^{\alpha\beta\gamma} A^{\gamma}_{i} W^{\beta}_{\mu}(x) - \partial_{\mu} \Lambda^{\alpha}(x)\, A^{\alpha}_{i} ,$$

$$\delta W^{\alpha}_{\mu}(x) = \Lambda^{\gamma} f^{\gamma\alpha\beta} W^{\beta}_{\mu}(x) - \partial_{\mu} \Lambda^{\alpha}(x) .$$

Thus W^{α}_{μ} transforms as gauge bosons (with "coupling constant" -1). In the same way, we get the right transformation for $g_{\mu\nu}$ and g_{ij}. We shall write, for the gauge bosons:

$$B^{\alpha}_{\mu} = (h^{\alpha\beta} - \delta^{\alpha\beta})\, W^{\beta}_{\mu} .$$

This ansatz is dictated by the same feature $B^{\alpha}_{i} = A^{\alpha}_{i}$ as above.

Now, let us investigate the case $D = 1$. S is a circle. We need no matter fields. Thus, consider the metric:

$$g = \begin{bmatrix} g_{\mu\nu} + A_{\mu} A_{\nu} & \varphi^{2} A_{\mu} \\ \varphi^{2} A_{\mu} & \varphi^{2} \end{bmatrix} .$$

Nothing depends on the fifth coordinate θ (but φ depends on x^{μ}). For the Ricci tensor $R^{\hat{\mu}}{}_{\hat{\nu}}$ we get:

$$R^{5}{}_{5} = \frac{\Box\varphi}{\varphi} - \frac{1}{4} \varphi^{2} F_{\mu\nu} F^{\mu\nu} - A_{\mu} R^{\mu}{}_{5} ,$$

$$R^{\mu}{}_{5} = \tfrac{1}{2} D_{\nu} \varphi^{3} F^{\nu\mu} ,$$

$$R^{\mu}{}_{\nu} = R^{0\mu}_{\nu} + \frac{1}{\varphi} D^{\mu} D_{\nu} \varphi + \tfrac{1}{2} \varphi^{2} F^{\mu\lambda} F_{\nu\lambda} + A_{\nu} R^{\mu}{}_{5} ,$$

$$R = R^{0} + \tfrac{1}{4} \varphi^{2} F^{\mu\nu} F_{\mu\nu} + \frac{2\Box\varphi}{\varphi} ,$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $R^{0\mu}_{\ \ \nu}$ is the curvature calculated with $g_{\mu\nu}$. (This is the result of Kaluza-Klein, for φ = constant.)

Thus, if we set φ = constant, the Einstein equations lead to $F_{\mu\nu}F^{\mu\nu} = 0$.

This indicates that, in the general case, there are scalars, coupled with the gauge fields. Thus, the ansatz given above cannot be a solution of the equations of motion (without quadratic constraints), and is only useful to get the expansion of the Lagrangian in a series of "harmonics".

We shall give the main results of the computation.

First, one checks that the mass of W^α_μ vanishes. After that, we compute the Ricci tensor. The easiest way is to use the "vierbein" formalism, writing:

$$g_{\hat\mu\hat\nu} = e^{\hat a}{}_{\hat\mu}\, e^{\hat b}{}_{\hat\nu}\, \eta_{\hat a\hat b}\ ,$$

where $\eta_{\hat a\hat b}$ is the flat metric $(^{-1}_{\ 0}, +1\ ^{\ 0}_{+1})$ we must calculate the connection $\omega^{\hat a}{}_{\hat b\hat\mu}$ such that:

$$\omega_{\hat a\hat b\hat\mu} + \omega_{\hat b\hat a\hat\mu} = 0\ ,$$

$$\omega^{\hat a}{}_{\hat b\hat\mu}\, e^{\hat b}{}_{\hat\nu} - \omega^{\hat a}{}_{\hat b\hat\nu}\, e^{\hat b}{}_{\hat\mu} + \partial_{\hat\mu}\, e^{\hat a}{}_{\hat\nu} - \partial_{\hat\nu}\, e^{\hat a}{}_{\hat\mu} = 0\ .$$

Here,

$$e^{\hat a}{}_{\hat\mu} = \begin{bmatrix} e^\lambda{}_\mu & 0 \\ e^i{}_j A^{j\alpha} W^\alpha_\mu & e^i{}_j \end{bmatrix} \quad \begin{matrix} e^\lambda{}_\mu = \text{4-dimensional vierbein} & g_{\mu\nu} = \eta_{\lambda\rho}\, e^\lambda{}_\mu e^\rho{}_\nu \\ e^i{}_j = D\text{-dimensional vierbein} & g_{ij} = \delta_{kl}\, e^k{}_i e^l{}_j \end{matrix}$$

we define:

$$F^\alpha_{\mu\nu} = \partial_\mu W^\alpha_\nu - \partial_\nu W^\alpha_\mu - f^{\alpha\beta\gamma} W^\beta_\mu W^\gamma_\nu\ ,$$

then

$$\omega_{ijk} = \omega^0_{ijk} \qquad (D\text{-dimensional connection})\ ,$$

$$\omega_{i\mu j} = 0\ ,$$

$$\omega_{ij\mu} = -\tfrac{1}{2} e_i{}^k e_j{}^l (\partial_k A^\alpha_l - \partial_l A^\alpha_k)\, W^\alpha_\mu\ , \qquad e_i{}^k \text{ is the inverse of } e^i{}_k\ ,$$

$$\omega_{i\nu\mu} = \tfrac{1}{2} e_\nu{}^\lambda\, e_i{}^k A^\alpha_k F^\alpha_{\lambda\mu}\ ,$$

$$\omega_{\mu\nu i} = -\tfrac{1}{2} e_\mu{}^\lambda\, e_\nu{}^\rho A^\alpha_i F^\alpha_{\lambda\rho}\ ,$$

$$\omega_{\mu\nu\rho} = \omega^0_{\mu\nu\rho} - \tfrac{1}{2} W^\alpha_\rho e_\mu{}^\lambda\, e_\nu{}^\sigma F^{\alpha'}_{\lambda\sigma} h^{\alpha\alpha'}\ , \qquad \omega^0_{\mu\nu\rho} = \text{4-dimensional connection.}$$

(recall $h^{\alpha\alpha'} = A^{i\alpha} A^{\alpha'}_i$.)

Then the curvature $R^{\hat{a}}{}_{\hat{b}\hat{\mu}\hat{\nu}}$ is given by:

$$R^{\hat{a}}{}_{\hat{b}\hat{\mu}\hat{\nu}} = \partial_{\hat{\mu}}\omega^{\hat{a}}{}_{\hat{b}\hat{\nu}} - \partial_{\hat{\nu}}\omega^{\hat{a}}{}_{\hat{b}\hat{\mu}} + \omega^{\hat{a}}{}_{\hat{c}\hat{\mu}}\omega^{\hat{c}}{}_{\hat{b}\hat{\nu}} - \omega^{\hat{a}}{}_{\hat{c}\hat{\nu}}\omega^{\hat{c}}{}_{\hat{b}\hat{\mu}} \,.$$

Furthermore $R_{\hat{b}\hat{\mu}} = e_{\hat{a}}{}^{\hat{\nu}} R^{\hat{a}}{}_{\hat{b}\hat{\mu}\hat{\nu}}$ and $R_{\hat{\mu}\hat{\nu}} = e_{\hat{\mu}}{}^{\hat{b}} R_{\hat{b}\hat{\nu}}$. For the Ricci tensor with $\hat{\mu}$ indices, we get:

$$R^{i}{}_{j} = R^{0i}{}_{j} - \tfrac{1}{4}A^{\alpha}_{i}A^{\beta}_{j}F^{\mu\nu\alpha}F^{\beta}_{\mu\nu} - A^{i\alpha}W^{\alpha}_{\mu}R^{\mu}{}_{j} \,,$$

$$R^{\mu}{}_{i} = \tfrac{1}{2}(\hat{D}_{\nu}F^{\nu\mu\alpha})\,A^{\alpha}_{i} \,,$$

$$R^{\mu}{}_{\nu} = R^{0\mu}{}_{\nu} + \tfrac{1}{2}h^{\alpha\beta}F^{\mu\lambda\alpha}F^{\beta}_{\nu\lambda} + A^{i\alpha}W^{\alpha}_{\nu}R^{\mu}{}_{i} \,,$$

$$R = R^{0} + R^{S} + \tfrac{1}{4}h^{\alpha\beta}F^{\alpha}_{\mu\nu}F^{\mu\nu\beta} \,,$$

R^0 is the 4-dimensional curvature and R^S = curvature of S.
Furthermore $T^{\mu}_{i} = 0$. (This indicates that the Yang-Mills equation holds when scalars are not taken into account.)

We can also calculate the action for this ansatz.

We take the contribution given by $-(1/4e^2)\,G^{\alpha}_{\hat{\mu}\hat{\nu}}G^{\alpha\hat{\mu}\hat{\nu}}$. It is

$$-\frac{1}{4e^2}\,G^{\alpha}_{ij}G^{\alpha ij} - \frac{1}{2e^2}(\delta^{\alpha\beta} - h^{\alpha\beta})\,F^{\alpha}_{\mu\nu}F^{\beta\mu\nu} \,.$$

Remarking that the Lagrangian must be a scalar for the group G (because G acts as a general coordinate transformation, and a gauge transformation) the second term is easy to obtain.

Furthermore

$$\sqrt{-g} = \sqrt{-g^0}\,\sqrt{g^S} \,,$$

$$g^0 = \det|g_{\mu\nu}| \,,$$

$$g^S = \det|g_{ij}| \,.$$

At zero order in W^{α}_{μ}, the action vanishes, as mentioned above. Then

$$S = -\frac{1}{16\pi G}\int d^4x d^D\theta\,\sqrt{-g^0}\sqrt{g^S}(R^0 + \tfrac{1}{4}\mu\, h^{\alpha\beta}F^{\mu\nu\alpha}F^{\beta}_{\mu\nu})$$

$$-\frac{1}{2e^2}\int d^4x d^D\theta\sqrt{-g^0}\sqrt{g^S}(\delta^{\alpha\beta} - h^{\alpha\beta})\,F^{\alpha}_{\mu\nu}F^{\mu\nu\beta} \,,$$

$$(\mu = 8\pi G/e^2) \,.$$

We can do the integration over the internal coordinates:

$$\int d^D\theta\,\sqrt{g^S}\,h^{\alpha\beta} = \epsilon\delta^{\alpha\beta} \,,$$

(because the results must be invariant under G). Because $h^{\alpha\alpha} = D$ we get $\epsilon = V_S D/N$ (V_S = volume of S, N = dim G.). Then

$$S = -\frac{1}{16\pi G^{(4)}}\int d^4xR^0\sqrt{-g^0} - \frac{1}{4(e^{(4)})^2}\int d^4x\sqrt{-g^0}F^{\alpha}_{\mu\nu}F^{\mu\nu\alpha},$$

with

$$G^{(4)} = \frac{G}{V_S},$$

$$(e^{(4)})^2 = \frac{e^2}{V_8\eta},$$

$$\eta = 2 - \frac{3}{2}\frac{D}{N}.$$

Remark that $G^{(4)}/(e^{(4)})^2 \sim G/e^2$. The value of μ can be computed in terms of $G^{(4)}$ and $e^{(4)}$ (only $G^{(4)}$, $e^{(4)}$ are experimentally observable). So, if we take Newton's constant for $G^{(4)}$ and a coupling constant of order α for ${e^{(4)}}^2$, we get the result that the size of the internal is always of the order of Planck's length, i.e. 10^{-33} cm. So, the 4-dimensional Lagrangian one obtains by expanding around our classical solution, is just that of gravitation and of a gauge theory. What happens essentially is that our solution breaks gauge invariance in $(4 + D)$ dimensions, but 4-dimensional gauge invariance remains unbroken. But, as mentioned above, the ansatz is not a solution of the equations of motion: so the expansion of the Lagrangian is incomplete. In order to find a better expansion, we shall forget the high-mass states of spin 1 and 2, and keep only scalars. This is not crucial for the following.
So, we set

$$g_{\hat{\mu}\hat{\nu}} = \begin{bmatrix} g_{\mu\nu}(x, \lambda) + g_{\mu i}g_{\nu j}g^{ij} & g_{\mu j}(x, \theta) \\ g_{i\nu}(x, \theta) & g_{ij}(x, \theta) \end{bmatrix},$$

$g_{ij}(x, \theta)$ is arbitrary,

$$g_{\mu i}(x, \theta) = g_{ij}(x, \theta)\, A^{j\alpha}(\theta)\, W^{\alpha}_{\mu}(x).$$

As before, one can check that W^{α}_{μ} transforms again as a gauge field, for the same action of G as above. $g_{ij}(x, \theta)$ will represent an infinite set of scalars. For example, one can write

$$g^{ij}(x, \theta) = A^{i\alpha}(\theta)\, A^{\delta\beta}(\theta)\, M_{\alpha\beta}(x, \theta).$$

Furthermore, we set for the gauge fields

$$B^{\alpha}_{\mu} = (A^{i\alpha}B^{\beta}_{i}\delta^{\alpha\beta})\, W^{\beta}_{\mu}(x),$$

B_i^α arbitrary, e.g.

$$B_i^\alpha = L^{\alpha\beta} A_i^\beta \ .$$

We call the scalars collectively Φ. Because of the gauge invariance the expansion of the Lagrangian will take the form

$$\mathcal{L}^{(4)} = -\frac{\sqrt{g^0}}{16\pi G^{(4)}} R^0 \alpha(\Phi) - \frac{\sqrt{-g^0}}{4(e^{(4)})^2} \beta(\Phi) F^\alpha_{\mu\nu} F^{\mu\nu\alpha} - \tfrac{1}{2}\gamma(\Phi)(\hat{D}_\mu \Phi \, \hat{D}^\mu \Phi) + \dots ,$$

where $F^\alpha_{\mu\nu}$ is the same as before, and α, β, γ are scalar functions such that

$$\alpha(0) = \beta(0) = 1 \ .$$

Note that we can eliminate $\alpha(\Phi)$ (variation of the constant of gravitation) by a conformal transformation on $g_{\mu\nu}$.

We don't know whether there are massless scalars. If not, we can forget them, and the first expansion is sufficient at usual energies. If there are massless scalars, and with the likely hypothesis that there are no other massless spin-1 states (except the massless spin-1 states we shall find later), the observation at four dimensions is: gravitation + gauge fields + massless scalars coupled in a covariant way. We have an indication that there are no massless scalars: for $D = 1$, the radius of the circle is a massless scalar, because Einstein's equation does not fix its value. For $D > 1$, the curvature is fixed by V_0. So, it seems to us that the apparition of such scalars would be accidental.

4. Symmetry breaking

Until now, we have assumed that K is G. More generally, we shall consider the case when G ≠ K. So, we choose a subgroup G of K, and normalize the structure constants of K such that: for $\alpha, \beta, \gamma, \delta$ corresponding to G,

$$f^{\alpha\beta\gamma} f^{\delta\rho\gamma} = \delta^{\alpha\delta} \ .$$

Thus, we can find a solution of the equations described in sect. 2 by setting:

$$B^{i\alpha} = \lambda A^{i\alpha} \qquad \text{in the direction of G} ,$$

$$B^{i\alpha} = 0 \qquad \text{otherwise} . \tag{9}$$

(Then the value $\mu = 8\pi G/e^2$ changes a little.)

After that, as in sect. 3 we get a gauge theory of G. But if G′ is a maximal subgroup of K commuting with G, it will also be left unbroken. Thus, we get a gauge theory: G × G′. We shall now inspect the spectrum of spin-1 fields (besides those already obtained from G) obtained by expanding the B_μ^α fields

$g_{\hat{\mu}\hat{\nu}}$ is not changed;
B_i^α is not changed (same as (9));

B^{α}_{μ} is changed (it must decouple from the linearized Einstein equations, for consistency).

The linearized Yang-Mills equations, for B^{α}_{μ}, are:

$$D_{\mu}G^{(1)\mu\nu\alpha} + D_i G^{(1)i\nu\alpha} + f^{\alpha\beta\gamma}B^{\beta}_{i}G^{(1)i\nu\gamma} = 0 \,,$$

$$D_{\mu}G^{(1)\mu i\alpha} = 0 \,,$$

$$G^{(1)\alpha}_{\mu\nu} = \partial_{\mu}B^{\alpha}_{\nu} - \partial_{\nu}B^{\alpha}_{\mu} \,, \qquad G^{(1)\alpha}_{i\nu} = \partial_i B^{\alpha}_{\nu} + f^{\alpha\beta\gamma}B^{\beta}_{i}B^{\gamma}_{\nu} \,.$$

So, fluctuations such that: $G^{\alpha}_{i\nu} = 0$ will be massless, and will, at first order, decouple from the Einstein equations (because $T^{(1)}_{\mu i} = 0$).

A trivial solution is obtained if α corresponds to G′ and we get massless vector bosons associated with the G′ subgroup of K. However, non-trivial solutions may exist: let us set

$$B^{\alpha}_{\mu}(x,\theta) = U_{\mu}(x)\,\varphi^{\alpha}(\theta) \,.$$

So we must search for scalar functions $\varphi^{\alpha}(\theta)$ on S, such that

$$\partial_i\varphi^{\alpha} + \lambda f^{\alpha\beta\gamma}A^{\beta}_{i}\varphi^{\gamma} = 0 \,. \tag{10}$$

The Lie algebra of G acts on that of K, by the relation

$$\mathrm{ad}\,X\cdot Y = [X, Y] \,.$$

We can split the Lie algebra of K into irreducible representations R^a. If we denote by T^{α} the generators for the action of G in R^a, (10) takes the form

$$\partial_i\varphi^{\alpha} + \lambda A^{\beta}_{i}(T^{\beta})^{\alpha}{}_{\gamma}\,\varphi^{\gamma} = 0 \,.$$

Furthermore: $L^{\alpha} = A^{i\alpha}\partial_i$ (differential operators on S), and

$$\partial_i = \frac{1}{\mu}A^{\alpha}_{i}L^{\alpha} \,.$$

Then, in each representation R^a (recall $\lambda\mu = 1$ for the type-II spaces that we consider) eq. (10) becomes symbolically

$$A^{\alpha}_{i}(L^{\alpha} + T^{\alpha})\,\varphi = 0 \,.$$

The scalar function on S gives representations S^b of G. Eq. (11) means that we must find a representation $S^b = R^a$. If this is the case, one obtains one vector $U^{a}_{\mu}(x)$ for the representation R^a.

In particular, the Lie algebra of G' gives several trivial representations, and the corresponding fields obviously form a gauge field of G'. But other massless bosons can appear. For them, the topological properties of S are important, as they can suppress the representations of S^b.

Note that the invariance of G′ is more than a four-dimensional gauge invariance:

this leads, for example, to massless scalar bosons, in the components B_i^α, for α in the "direction" of G′.

This generalizes to 4 + D dimensions the results obtained in [8] where the full mass spectrum of the broken generators was explicitly computed.

Let us give now an example.

The dual spinor model predicts $D = 6$. We can realize this space by:

$$S = S_1 \times S_2 \qquad S_1 = \frac{SU(3)}{SU(2) \times U(1)}, \qquad \text{dim } 4\,,$$

$$S_2 = \frac{SU(2)}{U(1)}, \qquad \text{dim } 2 \text{ (sphere)}\,.$$

S_1 and S_2 are spaces of type II. We can find a solution with gauge group K = SU(5), G = SU(3) × SU(2). Then G′ = U(1).

The massless bosons are obtained from gauge fields by expanding the Lie algebra of SU(5) with respect to SU(3) × SU(2):

$$24 = 8 \otimes 1 \oplus 1 \otimes 3 \oplus 1 \otimes 1 \oplus 3 \times 2 \oplus \bar{3} \times 2;$$

$1 \otimes 1$ give the gauge boson of G′ = U(1);

$3 \otimes 2$ and $\bar{3} \otimes 2$ are not obtained on the space $\delta_1 \times \delta_2$;

$8 \otimes 1$ and $1 \otimes 3$ give two massless vectors, and an $U(1)^2$ invariance.

So we get a gauge group: $SU(3) \times SU(2) \times U(1)^3$. The other bosons acquire superheavy masses.

So, for instance, starting from the Georgi-Glashow [14] model with K = SU(5), we obtain an unbroken gauge invariance which is $SU(3) \times SU(2) \times U(1)^3$. Leptoquark bosons naturally acquire a huge mass. On the other hand many questions remain to be solved before such a model can be considered to be realistic.

(i) We must look at the fermion spectrum (this has so far been done only for $D = 2$ (see ref. [8]) in order to see if the flavour could appear as a degeneracy of massless multiplets.

(ii) The only mass scale in this model is Planck's mass. Can one introduce other, much smaller mass scales, like the W boson mass, by considering solutions which are not exactly, but "almost" invariant under G? More generally, can one adjust exactly the space-time geometry of the classical solution to the pattern of broken symmetries observed in Nature, thus replacing the Higgs mechanism by a geometric symmetry breaking?

I am grateful to E. Cremmer and J. Scherk for their constant help during this work.

Note added

After completion of our work we learned that gauge fields on coset spaces have been considered also by G. Domokos and S. Kövesi-Domokos [15] with the purpose of reducing the number of independent gauge fields.

References

[1] Th. Kaluza, Sitzungber. Preuss. Akad. Wiss. Berlin, Math. Phys. K1, 966 (1921);
O. Klein, Z. Phys. 37 (1926) 895;
J. Rayski, Acta Phys. Pol. 27 (1965) 89.
[2] R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
N.P. Konoplyova and V.N. Popov, The gauge fields (Atomizdat, Moscow, 1972) (in Russian);
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
Y.M. Cho and Pong Soo Jang, Phys. Rev. D12 (1975) 3789;
J. Rayski, Dublin preprint DIAS-TP-76-31, to be published.
[3] A. Neveu and J.H. Schwarz, Nucl. Phys. B31 (1971) 86; Phys. Rev. D4 (1971) 1109;
P. Ramond, Phys. Rev. D3 (1971) 2415.
[4] F. Gliozzi, J. Scherk and D. Olive, Phys. Lett. 65B (1976) 282; Nucl. Phys. B122 (1977) 253.
[5] H. Fritzsch and P. Minkowsky, Ann. of Phys. 93 (1975) 193.
[6] E. Gildener, Phys. Rev. D14 (1976) 1667.
[7] H. Georgi, H.R. Quinn and S. Weinberg, Phys. Rev. Lett. 33 (1974) 451.
[8] Z. Horvath, L. Palla, E. Cremmer and J. Scherk, Nucl. Phys. B127 (1977) 57.
[9] E. Cremmer and J. Scherk, Nucl. Phys. B108 (1976) 409; B118 (1977) 61.
[10] D.Z. Freedman, P. Van Nieuwenhuizen and S. Ferrara, Phys. Rev. D13 (1976) 3214;
S. Deser and B. Zumino, Phys. Lett. B62 (1976) 335;
S. Ferrara and P. Van Nieuwenhuizen, Phys. Rev. Lett. 37 (1976);
S. Ferrara, J. Scherk and B. Zumino, Phys. Lett. 66B (1977) 35; Nucl. Phys. B121 (1977) 393;
D.Z. Freedman, Phys. Rev. Lett. 38 (1976) 105;
D.Z. Freedman and A. Das, Nucl. Phys. B120 (1977) 221;
A. Das, ITP-SB-77-4 preprint, to be published;
E. Cremmer, J. Scherk and S. Ferrara, LPTENS 77/7, to be published.
[11] S. Helgason, Differential geometry and symmetric spaces (Academic Press, New York and London, 1962).
[12] G. 't Hooft, Nucl. Phys. B79 (1974) 276.
[13] E. Cremmer and J. Scherk, Nucl. Phys. B103 (1976) 399.
[14] M. Georgi and S.L. Glashow, Phys. Rev. Lett. 32 (1974) 438.
[15] G. Domokos and S. Kövesi-Domokos, Desy preprint 77/08, to be published; Preprint JHU HET 774 (April 1977) revised version.

On Kaluza–Klein Theory

ABDUS SALAM

International Centre for Theoretical Physics, Trieste, Italy, and
Imperial College, London, England

AND

J. STRATHDEE

International Centre for Theoretical Physics, Trieste, Italy

Received December 8, 1981

Assuming the compactification of $4+K$-dimensional space-time implied in Kaluza–Klein-type theories, we consider the case in which the internal manifold is a quotient space, G/H. We develop normal mode expansions on the internal manifold and show that the conventional gravitational plus Yang–Mills theory (realizing local G symmetry) is obtained in the leading approximation. The higher terms in the expansions give rise to field theories of massive particles. In particular, for the original Kaluza–Klein $4+1$-dimensional theory, the higher excitations describe massive, charged, purely spin-2 particles. These belong to infinite dimensional representations of an $O(1,2)$.

I. INTRODUCTION

The motivation for studying generally covariant field theories in space-time of more than four dimensions is to obtain a geometrical interpretation of internal quantum numbers such as electric charge, i.e., to place them in the same context as energy and momentum [1, 2]. The latter observables are associated with translational symmetry in 4-dimensional Minkowski space; the internal observables would be associated with symmetry motions in the extra dimensions.

In theories of the Kaluza–Klein type one starts with the hypothesis that spacetime has $4+K$ dimensions. One assumes general covariance and adopts the $4+K$-dimensional curvature scalar as the Lagrangian. Next, one supposes that because of some dynamical mechanism, the ground state of this system is partially compactified [3], $M^4 \times B^K$ rather than M^{4+K}, where M^4 denotes 4-dimensional Minkowski space and B^K is a compact K-dimensional space [4]. The size of B^K must be sufficiently small to render it unresolvable at the currently available energies. For example, with the size of $B^K < 10^{-17}$ cm, it would be invisible to probes of energy $<10^3$ GeV, this being an upper limit.

Models have been constructed which illustrate this phenomenon of spontaneous compactification [5, 6]. In general, it is necessary for the system to include non-geometrical (matter) fields, coupled to the metric field, for this effect to be triggered. The presence of such *ad hoc* matter fields detracts from the simplicity of the purely geometrical theory and perhaps indicates that the geometrical component is not truly fundamental. An interesting exception, however, could be the extended supergravities where certain matter-like fields are justified by an underlying supersymmetry [7] and possibly the related geometry of superspace.

Assuming that the compactification is achieved, what are the implications of taking the extra dimensions seriously? By expanding the fields in a Kaluza–Klein theory in terms of normal modes on the compact space B^K, the extra dimensions make their appearance as massive multiplets of the associated symmetry group. In this paper we wish to examine these multiplets and the higher massive excitations.

To illustrate the emergence of masses, consider the original theory of Kaluza and Klein which involves one new co-ordinate, y. The 5-dimensional line element was assumed to take form

$$ds^2 = (dx^m e_m{}^a(x))^2 - (dy - dx^m \kappa A_m(x))^2,$$

where $m, n, a, b, \ldots$ take the values 0, 1, 2, 3 and $e_m{}^a(x)$ represents the vierbein field on 4-dimensional spacetime. The 4-vector $A_m(x)$ is to be interpreted as the electromagnetic potential. Gauge transformations here take the form of coordinate transformations

$$x^m \to x^m,$$

$$y \to y + \Lambda(x),$$

with

$$e_m{}^a \to e_m{}^a,$$

$$A_m \to A_m + \frac{1}{\kappa}\partial_m \Lambda,$$

which leave the 1-forms, $dx^m e_m{}^a$ and $dy - dx^m \kappa A_m$, invariant. The invariant coupling to a complex scalar field, $\phi(x, y)$, would be described by the Lagrangian

$$\det e[|e_a{}^m(\partial_m \phi + \kappa A_m \partial_y \phi)|^2 - |\partial_y \phi|^2 - \mu^2 |\phi|^2].$$

If one suppose that the y-dependence of ϕ is such that

$$\partial_y \phi = -\frac{i}{R}\phi$$

then this Lagrangian reduces to that for a charged scalar,

$$\det e\left[|e_a{}^m(\partial_m \phi - ieA_m \phi)|^2 - \left(\mu^2 + \frac{1}{R^2}\right)|\phi|^2\right],$$

where the charge is given by

$$e = \frac{\kappa}{R}.$$

The electric charge is hereby understood in terms of the radius, R, of the extra dimension, interpreted as a circle. Conversely, given e, this equation fixes R (the size of the manifold) as κ/e where κ is Planck length. This also implies that the masses are in Planck units $\sim\kappa^{-1} \sim 10^{19}$ GeV. Because of this one is naturally interested to discover any exceptional cases where, for reasons of symmetry, the mass contribution is suppressed. An example is the vector A_m itself which describes a massless particle, the photon. The gauge symmetry here associated with the transformation $y \to y + \Lambda(x)$ prevents it from acquiring any mass.

To consider non-Abelian symmetries one has to interpret the new co-ordinates y^μ, $\mu = 1, 2, \ldots, \kappa$ as a parametrization of the manifold of a compact non-Abelian group, G. The line element is now assumed to take the form [4]

$$ds^2 = (dx^m e_m{}^a(x))^2 - ((dy^\mu - dx^m \kappa A_m{}^\alpha(x)\, K_\alpha{}^\mu(y))\, e_\mu{}^\beta(y))^2,$$

where the auxiliary functions $e_\mu{}^\beta(y)$ constitute an orthonormal K-bein on the group manifold G, and the functions $K_\alpha{}^\mu(y)$ are Killing vectors on G. This formalism will be reviewed in some detail in the following sections.

A further generalization is possible where the new co-ordinates are taken to parametrize some homogeneous space, G/H, on which G can act [8, 9]. A novel aspect of this kind of scheme concerns the problem of embedding. In order to give a G-invariant meaning to the ground state it is necessary to associate the motions of G/H with frame rotations (or tangent space transformations). This requires that the stability group, H, be embedded in the tangent space group, $SO(K)$, in such a way that the K-vector of $SO(K)$ has the H-content of G/H.

In all these theories the metric field carries an infinite number of new degrees of freedom corresponding to the propagation of excitations in the new dimensions. As a rule these excitations are ignored because of their large masses. However, the massive excitations are not always ignorable: their contributions to a typical scattering amplitude are comparable to those due to graviton exchange. To see this, consider the scattering of charged scalar particles with the interactions represented schematically by

$$\mathscr{L}_{\text{int}} \sim \kappa h_{mn} \partial_m \phi^* \partial_n \phi + ieA_m \phi^* \overleftrightarrow{\partial}_m \phi + \cdots,$$

where h_{mn} denotes the graviton field and A_m one of the massive excitations. The amplitude for graviton exchange is of order $\kappa^2 E^2$, while that for A-exchange is of order $e^2 E^2/M^2$, where E represents the energy scale of the scattering process. Since $M \sim e/\kappa$ these contributions are of the same order.[1] Only in the region of low momentum transfer, $t \ll E^2$, does the graviton amplitude, $\sim\kappa^2 E^4/t$, become dominant.

---

[1] This comparability of contributions may have relevance to the problem of renormalizability of Kaluza–Klein theories.

The purpose of this study is to deal with some of the general features of Kaluza–Klein theories. These include the geometry of homogeneous spaces, the transformations of fields defined over them, the structure of normal mode expansions[2] on G/H, the embedding of Yang–Mills gauge transformations in the group of general co-ordinate transformations, and the extraction of an effective 4-dimensional Lagrangian. Such an effective action will be particularly useful for low energy effects ascribable to the existence of the microscopic and as yet unresolvable structure on space-time. Since much of the recent work in this field has been couched in the terminology of fibre bundles [10] it is not accessible to those without expertise in such matters. We shall therefore adopt a less sophisticated approach and express results in the more generally familiar component notation.

The plan of the paper is as follows. Some definitions, notational conventions and a brief description of the $4+K$-dimensional Einstein theory are given in Section II. There follows in Section III a discussion of symmetry expects of the compactification of the extra K dimensions. Harmonic expansions on G/H are introduced to describe the excitations. An important ingredient here is the embedding of the stability group, H, in the tangent space group, $SO(K)$. Integration over G/H is used to reduce the Lagrangian to effective 4-dimensional form. Some of this is illustrated in Section IV where the low energy sector of the $4+K$-dimensional theory is shown to contain the 4-dimensional Einstein and Yang–Mills terms. Much of the pedagogical detail is confined to the Appendices, except for Appendix V which deals with the spectrum of $4+1$ Kaluza–Klein theory.

II. EXTENDED GRAVITY

Let the $4+K$-dimensional spacetime be parametrized, at least in some local region, by a set of co-ordinates, z^M. On this space let there be defined at each point a local frame of reference in the form of $4+K$ linearly independent covariant vectors, ${E_M}^A$.

The basic symmetries of the theory are twofold. Firstly, there must be general covariance under the coordinate transformations, $z^M \to z'^M$. Being covariant vectors, the ${E_M}^A$ transform according to

$$ {E_M}^A(z) \to E'^A_M(z') = \frac{\partial z^N}{\partial z'^M}\, {E_N}^A(z). \tag{2.1}$$

Secondly, there is to be covariance under a group of z-dependent linear transformations among the vectors of the reference frame,

$$ {E_M}^A(z) \to E'^A_M(z) = {E_M}^B(z)\, {a_B}^A(z). \tag{2.2}$$

---

[2] Throughout this paper B^K is assumed to be a quotient space. One may eventually contemplate internal manifolds (associated with symmetry breaking) which are less symmetrical.

In order to establish a pseudo-Riemannian geometry on the $4+K$-space it is usual to assume that these transformations belong to the pseudo-orthogonal group, $SO(1, 3+K)$. We shall adhere to this choice.

With frame rotations belonging to $SO(1, 3+K)$ one is led to consider the frame-independent combinations,

$$g_{MN} = E_M{}^A E_N{}^B \eta_{AB}, \tag{2.3}$$

which define the covariant components of the metric tensor. By an appropriate choice of basis, the $SO(1, 3+K)$ invariant tensor, η_{AB}, can be represented by the diagonal matrix,

$$\eta_{AB} = \mathrm{diag}(1, -1, -1, \ldots, -1). \tag{2.4}$$

Indices are raised and lowered in the usual way. The contravariant quantities $E_A{}^M$ and g^{MN} are defined as matrix inverses of $E_M{}^A$ and g_{MN}, respectively. (We employ the alphabetic convention that letters $M, N, \ldots$ from the middle alphabet will be world indices while those from the early alphabet, $A, B, \ldots$ shall be frame labels.)

The coupling of extended gravity to matter is conveniently illustrated by the case of the fundamental spinor of $SO(1, 3+K)$. This spinor, $\psi(z)$, has $2^{2+[K/2]}$ components where $[K/2]$ denotes the largest integer $\leqslant K/2$. Under frame rotations the components of ψ transform as an $SO(1, 3+K)$ spinor,

$$\psi(z) \to \psi'(z) = S(a^{-1})\, \psi(z), \tag{2.5}$$

but, under general co-ordinate transformations they are scalars,

$$\psi(z) \to \psi'(z') = \psi(z). \tag{2.6}$$

The partial derivatives $\partial_M \psi$ make a covariant world vector but they do not transform homogeneously under frame rotations. The latter defect is cured by introducing the spin connections $B_M(z)$, which constitute a gauge field with respect to frame rotations,

$$B_M \to S B_M S^{-1} + S \partial_M S^{-1} \tag{2.7}$$

and a covariant vector with respect to co-ordinate transformations. The covariant derivative is then defined by

$$\nabla_M \psi = (\partial_M + B_M)\, \psi. \tag{2.8}$$

The spin connections belong to the infinitesimal algebra of $SO(1, 3+K)$. In the spinor representation one can write

$$B_M = \tfrac{1}{8} B_{M[AB]} [\gamma^A, \gamma^B], \tag{2.9}$$

where γ^A is a generalized Dirac matrix, i.e.,

$$\{\gamma^A, \gamma^B\} = 2\eta^{AB}. \tag{2.10}$$

With these fields it is possible to construct a suitable Lagrangian density, one which is invariant under frame rotations and a scalar density under co-ordinate transformations

$$\mathscr{L}_\psi = \frac{i}{2} \det(E)\, \bar{\psi}\gamma^A E_A{}^M \nabla_M \psi + \text{h.c.} \tag{2.11}$$

This much is nothing but a straightforward generalization to $4 + K$ dimensions of the well-known 4-dimensional structures. However, when one comes to look for the Yang–Mills potentials among the components of E and B, one will find some less familiar things. For example, the "minimal coupling" of the vectors is contained partly in the term $\bar{\psi}\gamma^A E_A{}^M \partial_M \psi$ and partly in $\bar{\psi}\gamma^A E_A{}^M B_M \psi$. The latter term may also contain a non-minimal, Pauli type of coupling.

Out of E and B and their first derivatives it is possible to construct two distinct covariant objects, or "field strengths." These are the torsion,

$$T_{MN}{}^A = \partial_M E_N{}^A - \partial_N E_M{}^A + E_N{}^C B_{MC}{}^A - E_M{}^C B_{NC}{}^A, \tag{2.12}$$

and the curvature,

$$R_{MN[AB]} = \partial_M B_{N[AB]} - \partial_N B_{M[AB]} - B_{M[AC]} B_{N[CB]} + B_{N[AC]} B_M[CB] \tag{2.13}$$

(where early alphabet labels are raised, lowered and contracted with η_{AB} defined in (2.4)).

The curvature scalar (density) is given by

$$\det(E)\, R = \det(E)\, E_A{}^M E_B{}^N R_{MN[AB]} \tag{2.14}$$

and can serve as a Lagrangian density for the extended gravity fields. The fields E and B may be varied independently or, alternatively, one may impose the torsion constraints, $T_{MN}{}^A = 0$, and eliminate B as an independent variable. The constraints $T = 0$ are algebraic in B and can easily be solved. One finds

$$B_{M[BC]} = \tfrac{1}{2} E_M{}^A (\Omega_{[AB]C} - \Omega_{[BC]A} + \Omega_{[CA]B}), \tag{2.15}$$

where

$$\Omega_{[AB]C} = E_A{}^M E_B{}^N (\partial_M E_{NC} - \partial_N E_{MC}). \tag{2.16}$$

III. Compactification

Having set up an action principle in $4+K$ dimensions one might inquire into the nature of the solutions to the resulting equations of motion. In a purely geometrical theory one would expect the ground state, or vacuum, to exhibit maximal symmetry. Indeed, one should find the $4+K$-dimensional Minkowski space. However, in the presence of matter fields—scalars and vectors— the situation is more complicated and it is possible to imagine that the ground state will manifest some lower symmetry. In particular, if the theory is supersymmetric the fields which accompany E and B may trigger a compactification of some of the dimensions in the ground state. We shall assume that this happens and that the ground state geometry factorizes into the product of 4-dimensional Minkowski space with a compact homogeneous space, G/H, of K dmensions. This space is invariant under the action of some group G which will be interpreted as an "internal" symmetry. We shall further assume that the compact space is small enough (size $\sim\kappa$) to justify, for most purposes, the neglect of any excitations associated with it (i.e., that such excitations would be too massive to be accessible at present energies). As mentioned in the Introduction the excitations may not always be negligible. Their contributions are generally comparable to those of gravitons and they would certainly be important if the gravitons are quantized. The aim at this stage, however, is to develop an effective 4-dimensional theory for the long range degrees of freedom.

One way to obtain the 4-dimensional theory is to expand all field variables in complete sets of harmonics on the homogeneous space G/H. The expansion coefficients are 4-dimensional fields belonging to a sequence of representations of G. Their dynamics is governed by a 4-dimensional action functional obtained by integrating out the G/H-dependence and using the orthogonality properties of harmonics on G/H. The resulting masses and couplings are necessarily G-invariant. The scale of the masses is determined by the size of the compact manifold, and we shall assume that all non-vanishing masses are therefore large. One would like to be able to extract the zero modes and discard the rest. We shall consider some aspects of this problem in the following but do not obtain a general solution. We now develop the necessary formalism.

To distinguish the compact dimensions from the remaining four of ordinary spacetime it is convenient to adopt some notational conventions. For the co-ordinates write

$$z^M = (x^m, y^\mu) \tag{3.1}$$

with middle alphabet Latin suffices, $m, n,\ldots$ taking the values 0, 1, 2, 3 and their Greek counterparts, $\mu, \nu,\ldots$ taking the values $1, 2,\ldots K$. The compact manifold, G/H, is parametrized by y^μ. The early alphabet frame labels are likewise divided into lower case Latin and Greek.

To implement the parametrization of G/H there are available standard group-theoretical techniques [11], one version of which we outline briefly. Suppose the co-ordinates, y^μ, label the cosets of G with respect to the subgroup, H. That is, from each coset let there be chosen a representative element, L_y. Multiplication from the left by an arbitrary element $g \in G$ will generally carry L_y into another coset, one for which the representative element is $L_{y'}$. This defines the so-called left translation

$$gL_y = L_{y'}h, \tag{3.2}$$

where h is an element of H. Both y'^{μ} and h are determined by this equation as function of y^{μ} and g. The explicit form of these functions depends of course, on the choice of elements L_y which represent the cosets. Such details are not usually important and, here, we shall be able to manage with some general properties of the parametrization. Among these, the most important is the notion of a covariant basis on G/H.

To define a covariant basis consider the 1-form, or differential,

$$e(y) = L_y^{-1} dL_y. \tag{3.3}$$

This object belongs to the infinitesimal algebra of G and therefore can be expressed as a linear combination of the generators, $Q_{\hat{\alpha}}$,

$$\begin{aligned} e(y) &= e^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}} \\ &= dy^{\mu} e_{\mu}{}^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}}. \end{aligned} \tag{3.4}$$

The generators $Q_{\hat{\alpha}}$ fall in two categories: the set $Q_{\bar{\alpha}}$ which generates the subgroup, H, and the remainder, Q_{α}, $\alpha = 1, 2, \ldots, K$, associated with the cosets, G/H. Correspondingly, one writes

$$e(y) = e^{\alpha}(y)\, Q_{\alpha} + e^{\bar{\alpha}}(y)\, Q_{\bar{\alpha}}. \tag{3.5}$$

Of special interest are the coefficients $e_{\mu}{}^{\alpha}(y)$ which constitute a nonsingular $K \times K$ matrix. The columns of this matrix provide the covariant components of the K linearly independent vectors of the covariant basis. The contravariant components of the reciprocal basis, $e_{\alpha}{}^{\mu}$, are defined by inverting the matrix $e_{\mu}{}^{\alpha}$.

The behaviour of $e(y)$ under left translations can be deduced from (3.2) and (3.3). Thus, quite generally,

$$\begin{aligned} e(y) \to e(y') &= hL_y^{-1} g^{-1} d(gL_y h^{-1}) \\ &= he(y)\, h^{-1} + hdh^{-1} + hL_y^{-1} g^{-1} dgL_y h^{-1}. \end{aligned} \tag{3.6}$$

On substituting the expansion in terms of generators and introducing the matrices, $D_{\hat{\alpha}}{}^{\beta}$, of the adjoint representation of G, defined by

$$g^{-1} Q_{\hat{\alpha}}\, g = D_{\hat{\alpha}}{}^{\beta}(g)\, Q_{\beta} \tag{3.7}$$

the formula (3.6) gives

$$e^{\hat{\alpha}}(y') = e^{\beta}(y)\, D_{\beta}{}^{\hat{\alpha}}(h^{-1}) + (hdh^{-1})^{\hat{\alpha}} + (g^{-1}dg)^{\beta}\, D_{\beta}{}^{\hat{\alpha}}(L_y h^{-1}). \tag{3.8}$$

The components, $(hdh^{-1})^{\hat{\alpha}}$ and $(g^{-1}dg)^{\hat{\alpha}}$, of elements in the algebra are naturally defined by

$$\begin{aligned} g^{-1}dg &= (g^{-1}dg)^{\hat{\alpha}}\, Q_{\hat{\alpha}}, \\ hdh^{-1} &= (hdh^{-1})^{\hat{\alpha}}\, Q_{\hat{\alpha}} = (hdh^{-1})^{\bar{\alpha}}\, Q_{\bar{\alpha}}. \end{aligned} \tag{3.9}$$

Since hdh^{-1} belongs to the algebra of H, we must have $(hdh^{-1})^{\alpha} = 0$.

(A more detailed discussion of the properties of the "boost" elements, L_y, together with some illustrations is given in Appendix I. Curvature and torsion formulae for G/H are derived in Appendix II, and various implications of the transformation formula (3.8) are derived in Appendix III.)

Now consider the question of harmonic expansions on the internal space. To begin with, if the internal space were simply the group G itself (i.e., $H = 1$) it would be natural to employ the full set of matrices of the unitary irreducible representations of G. For the function, $\phi(g)$, one would write

$$\phi(g) = \sum_{n} \sum_{p,q} \sqrt{d^n}\, D^n_{pq}(g)\, \phi^n_{qp}. \tag{3.10}$$

where D^n_{pq} is a unitary matrix of dimension d_n and the sum includes all matrix elements of all the unitary irreducible representations, $g \to D^n(g)$. The coefficients, ϕ^m_{pq}, are projected out by integrating over the group,

$$\phi^n_{pq} = \frac{\sqrt{d_n}}{V_G} \int_G d\mu D^n_{pq}(g^{-1})\, \phi(g), \tag{3.11}$$

where $d\mu$ is the invariant measure normalized to volume V_G.

For functions on the coset space G/H the expansion is somewhat restricted. Typically, one is concerned with functions $\phi_i(g)$ that are subject to the auxiliary symmetry

$$\phi_i(hg) = \mathbb{D}_{ij}(h)\, \phi_j(g), \tag{3.12}$$

where $h \in H$ and $\mathbb{D}(h)$ is some particular representation of H. This means that $\phi(g)$, though not strictly constant over the points in a coset, is related by a linear rule, i.e.,

$$\phi_i(g_1) = \mathbb{D}_{ij}(g_1 g_2^{-1})\, \phi_j(g_2)$$

for g_1 and g_2 in the same coset. The appropriate restriction of the expansion (3.10) is clear. It must include only those terms for which

$$\mathbb{D}^n(hg) = \mathbb{D}(h)\, D^n(g).$$

For irreducible $\mathbb{D}(h)$ one should write

$$\phi_i(g) = \sum_{n} \sum_{\zeta,q} \sqrt{\frac{dn}{d_{\mathbb{D}}}}\, D^n_{i\zeta,q}(g)\, \phi^n_{q\zeta}, \tag{3.13}$$

where the sum includes all unitary irreducible representations of G for which $D^n_{i\zeta,q}$ makes sense, i.e., all those which include $\mathbb{D}(h)$ on restriction to H. If $\mathbb{D}(h)$ is contained more than once in a given $D^n(g)$ then a supplementary label, ζ, is needed to distinguish them. The dimension of $\mathbb{D}(h)$ is denoted $d_{\mathbb{D}}$. The coefficients, $\phi^n_{q\zeta}$, are projected from the sum (3.13) by integrating over the coset space,

$$\phi^n_{q\zeta} = \frac{1}{V_K}\sqrt{\frac{d_n}{d_{\mathbb{D}}}}\int_{G/H} d\mu D^n_{q,i\zeta}(L_y)\,\phi_i(L_y^{-1}). \tag{3.14}$$

This expression can be derived from (3.11) by resolving the general element g^{-1} into the product of appropriate elements L_y and h. Such a decomposition is unique once the form of the representative element L_y has been chosen. From (3.12) it follows that the h-dependence factorizes,

$$\begin{aligned}\phi_i(g) &= \phi_i(h^{-1}L_y{}^{-1})\\ &= \mathbb{D}_{ij}(h^{-1})\,\phi_j(L_y{}^{-1}).\end{aligned}$$

Integration over the subgroup H can be carried out explicitly and leaves the result (3.14).

Thus, for a set of fields $\psi_i(x,y)$ the general form of the harmonic expansion on G/H is given by

$$\psi_i(x,y) = \sum_n\sum_{\zeta,q}\sqrt{\frac{d_n}{d_{\mathbb{D}}}}\,D^n_{i\zeta,q}(L_y{}^{-1})\,\psi^n_{q\zeta}(x). \tag{3.15}$$

Here n labels the exitations.

The coefficients fields $\psi^n_{q\zeta}(x)$ are obtained by integrating over G/H as in (3.14). If the representation $\mathbb{D}(h)$ to which ψ_i belongs is reducible then it would be necessary to separate ψ_i into irreducible parts and make an expansion of the form (3.15) for each. This raises the important question of how the H content of $\psi(x,y)$ is to be determined. It will be settled by invoking the ground state symmetry.

The full set of fields in $4+K$ dimensions is supposed to be governed by an action functional which is invariant with respect to both the general co-ordinate transformations and the frame rotations. These symmetries will not be manifest in the ground state, however. By assumption the ground state geometry is $M^4\times G/H$, where G/H is a compact K-dimensional manifold. In the ground state we therefore have

$$\langle E_M{}^A(x,y)\rangle = \begin{pmatrix}\delta^a_m & 0\\ 0 & e_\mu{}^\alpha(y)\end{pmatrix}, \tag{3.16}$$

where $e_\mu{}^\alpha$ is the covariant basis on G/H defined above. The expression (3.16) is invariant under the product of the 4-dimensional Poincaré group and the local (x-dependent) transformations of G. The Poincaré invariance is exactly as in the familiar 4-dimensional theory of general relativity. The local G-invariance is compounded of the x-dependent left translations on G/H, $y\rightarrow y'(x,y)$ and the induced tangent space transformations, $D_\alpha{}^\beta(h)$, with h defined by (3.2). To see this, consider the action of the combined transformation on $E_\mu{}^\alpha(x, y)$

$$E_\mu{}^\alpha(x, y) \to E'_\mu{}^\alpha(x, y') = \frac{\partial y^\nu}{\partial y'^\mu} E_\nu{}^\beta(x, y)\, D_\beta{}^\alpha,$$

where $D_\beta{}^\alpha$ is a K-dimensional orthogonal matrix. In order to have

$$\langle E'_\mu{}^\alpha(x, y)\rangle = \langle E_\mu{}^\alpha(x, y)\rangle$$

it is only necessary to choose $\langle E_\mu{}^\alpha\rangle = e_\mu{}^\alpha$ and $D_\beta{}^\alpha = D_\beta{}^\alpha(h)$. This follows from (3.8) which contains the formula

$$dy'^\mu e_\mu{}^\alpha(y') = dy^\mu e_\mu{}^\beta(y)\, D_\beta{}^\alpha(h^{-1}).$$

Thus the embedding is fixed. With the left translation, $y \to y'$, must be associated the $SO(K)$ transformation, $D_\alpha{}^\beta(h)$, where $h \in H$. This implies that the K-vector of $SO(K)$ has the H-content of G/H. More precisely, the generators, $Q_{\bar\alpha}$, of H can be expressed in terms of the generators, $\Sigma^{\alpha\beta}$, of $SO(K)$ by the formula[3]

$$Q_{\bar\gamma} = -\tfrac{1}{2} c_{\bar\gamma\alpha\beta} \Sigma^{\alpha\beta}, \tag{3.17}$$

where the coefficients $c_{\bar\gamma\alpha\beta}$ are taken from the set of structure constants $c_{\hat\gamma\hat\alpha\beta}$ of G.

The Yang–Mills group of transformations described here, which leaves the ground state invariant also preserves the form of the expansion (3.15). Thus, with ψ transforming according to

$$\psi_i(x, y) \to \psi'_i(x, y') = \mathbb{D}_{ij}(h)\, \psi_j(x, y) \tag{3.18}$$

it is a simple matter to derive the corresponding rule for the expansion coefficients

$$\psi^n_{p\zeta}(x) \to \psi'^n_{p\zeta}(x) = D^n_{pq}(g)\, \psi^n_{q\zeta}(x). \tag{3.19}$$

They belong to the various irreducible representations of G which contribute to the expansion (3.15).

In order to express the full $4+K$-dimensional theory in its equivalent 4-dimensional form, one should substitute expansions like (3.15) for all of the fields in the system. The 4-dimensional Lagrangian is then obtained by integrating the $4+K$ dimensional one over the compact manifold, G/H. Integration of products of matrices

---

[3] For an infinitesimal transformation in H we have

$$\begin{aligned} D_{\alpha\beta}(h) &= \delta_{\alpha\beta} + \delta h^{\bar\gamma} c_{\alpha\bar\gamma\beta} \\ &= \delta_{\alpha\beta} + \omega_{\alpha\beta}, \end{aligned}$$

where $\omega_{\alpha\beta} = -\omega_{\beta\alpha}$. Viewed as an infinitesimal $SO(K)$ transformation this reads $(1/2)\,\omega_{\alpha\beta}\Sigma^{\alpha\beta} = \delta h^{\bar\gamma} Q_{\bar\gamma}$, which implies (3.17).

$D^n(L_y^{-1})$ reduces, in principle, to an algebraic problem. Although it is difficult to obtain results without going into the details of particular cases, the most important feature, orthogonality, can be illustrated by an example.

Consider the invariant integral

$$I^{nn'}_{p\zeta,p'\zeta'} = \int_{G/H} d^K y \det e(y)\, D^n_{p,i\zeta}(L_y)\, D^{n'}_{i\zeta',p'}(L_y{}^{-1}), \tag{3.20}$$

where the label, i, is understood to be summed. Since the measure, $d\mu = d^K y \det e$, is invariant under left translations, one can write

$$\begin{aligned} I^{nn'}_{p\zeta,p'\zeta'} &= \int d\mu D^n_{p,i\zeta}(L_{y'})\, D^n_{i\zeta',p'}(L_{y'}{}^{-1}) \\ &= \int d\mu D^n_{p,i\zeta}(gL_y h^{-1})\, D^n_{i\zeta',p'}(hL_y{}^{-1}g^{-1}) \\ &= D^n_{pq}(g)\, I^{nn'}_{q\zeta,q'\zeta'} D^{n'}_{q'p'}(g^{-1}) \end{aligned}$$

because the dependence on h cancels. Hence I is an invariant tensor of G. By Schur's lemma it is must take the form

$$I^{nn'}_{p\zeta,p'\zeta'} = \delta_{nn'}\delta_{pp'}k^n_{\zeta\zeta'}.$$

To evaluate $k^n_{\zeta\zeta'}$, set $n=n'$, $p=p'$ and sum over p. One obtains

$$\begin{aligned} d_n k^n_{\zeta\zeta'} &= \int d\mu D^n_{i\zeta',p}(L_y^{-1})\, D^n_{p,i\zeta}(L_y) \\ &= V_K d_{\mathrm{D}} \delta_{\zeta\zeta'}, \end{aligned}$$

where d_n and d_{D} are the dimensions of D^n_{pq} and $\mathbb{D}_{ij}$, respectively. Thus

$$I^{nn'}_{p\zeta,p'\zeta'} = V_K \frac{d_{\mathrm{D}}}{d_n} \delta_{\zeta\zeta'}\delta_{pp'}\delta_{nn'}. \tag{3.21}$$

This result is needed for inverting the expansion (3.15).

The argument leading to (3.21) generalizes. For example, one can write

$$\begin{aligned} &\int_{G/H} d\mu c^{i_1 i_2 i_3} D^{n_1}_{i_1\zeta_1,p_1}(L_y^{-1})\, D^{n_2}_{i_2\zeta_2,p_2}(L_y^{-1})\, D^{n_3}_{i_3\zeta_3,p_3}(L_y^{-1}) \\ &\quad = a_{n_1 n_2 n_3} V_K c^{i_1 i_2 i_3} \begin{pmatrix} n_1 & n_2 & n_3 \\ i_1\zeta_1 & i_2\zeta_2 & i_3\zeta_3 \end{pmatrix} \begin{pmatrix} n_1 & n_2 & n_3 \\ p_1 & p_2 & p_3 \end{pmatrix}, \end{aligned}$$

where $c^{i_1 i_2 i_3}$ is an invariant tensor of H and the 3–n symbols are essentially Clebsch–Gordan coefficients for the unitary representations of G. The factor $a_{n_1 n_2 n_3}$ depends on normalization conventions for the 3–n symbols.

Having given, in outline,[4] the general form of the expansions to be used, we shall in the following section restrict our considerations to the leading terms. The justification for this restriction is simply that, in the regime of low energies, only zero-modes are important.

IV. EFFECTIVE LAGRANGIANS

It is guaranteed that the graviton and a set of Yang–Mills vectors are included among the zero-mass fields of the 4-dimensional effective theory. This results from the gauge symmetries of the $4+K$-dimensional theory and the assumed symmetry of the ground state. It can be proved as follows. For $E_M{}^A(x,y)$ consider the ansatz [9]

$$E_M{}^A(x,y)=\begin{pmatrix} E_m{}^a(x) & -A_m{}^\beta(x)\,D_\beta{}^\alpha(L_y) \\ 0 & e_\mu{}^\alpha(y) \end{pmatrix}, \tag{4.1}$$

where the y-dependence is specified. This ansatz is compatible with a subgroup of the $4+K$-dimensional symmetries: the 4-dimensional general co-ordinate transformations,

$$x^m \to x'^m(x)$$

and the Yang–Mills transformations, or x-dependent left translations

$$y^\mu \to y'^\mu(x,y)$$

with associated frame rotations, $D_\alpha{}^\beta(h)$,

$$E'^a_m(x')=\frac{\partial x^n}{\partial x'^m}E_n{}^a(x), \tag{4.2a}$$

$$E'^\alpha_m(x',y')=\left(\frac{\partial x^n}{\partial x'^m}E_n{}^\beta(x,y)+\frac{\partial y^\mu}{\partial x'^m}e_\mu{}^\beta(y)\right)D_\beta{}^\alpha(h^{-1}), \tag{4.2b}$$

$$e_\mu{}^\alpha(y')=\frac{\partial y^\nu}{\partial y'^\mu}e_\nu{}^\beta(y)\,D_\beta{}^\alpha(h^{-1}). \tag{4.2c}$$

Of these rules, (4.2a) requires no comment and (4.2c) has been dealt with already in Section III. There remains (4.2b) which implies, in view of the ansatz,

$$-A'^\beta_m(x')\,D_\beta{}^\alpha(L_{y'})=\left(-\frac{\partial x^n}{\partial x'^m}A_n{}^\gamma(x)\,D_\gamma{}^\beta(L_y)+\frac{\partial y^\mu}{\partial x'^m}e_\mu{}^\beta(y)\right)D_\beta{}^\alpha(h^{-1}).$$

---

[4] Some illustrative examples are contained in Appendix IV.

With the help of formulae derived in Appx. 3 this rule reduces to the form

$$A'_m(x') = \frac{\partial x^n}{\partial x'^m}\,(gA_n(x)\,g^{-1} - g\partial_n g^{-1}), \tag{4.3}$$

where $A_m = A_m{}^{\hat{\alpha}} Q_{\hat{\alpha}}$. This is precisely the rule to be expected for a Yang–Mills potential.

The zero torsion spin connections corresponding to the ansatz (4.1) are obtained by substituting in the formulae (2.15), (2.16)

$$\begin{aligned}
B_{a[bc]} &= \tfrac{1}{2}E_{[a}{}^m E_{b]}{}^n\,\partial_m E_{nc} - \tfrac{1}{2}E_{[b}{}^m E_{c]}{}^n\,\partial_m E_{na} + \tfrac{1}{2}E_{[c}{}^m E_{a]}{}^n\,\partial_m E_{nb},\\
B_{a[b\gamma]} &= -B_{\gamma[ab]} = \tfrac{1}{2}E_a{}^m E_b{}^n F_{mn}{}^{\hat{\alpha}} D_{\hat{\alpha}\gamma}(L_y),\\
B_{a[\beta\gamma]} &= A_a{}^{\hat{\delta}}(D_{\hat{\delta}}{}^{\bar{\alpha}}(L_y) - D_{\hat{\delta}}{}^{\alpha}(L_y)\,\pi_\alpha{}^{\bar{\alpha}}(y))\,c_{\bar{\alpha}\beta\gamma},\\
B_{\alpha[\beta c]} &= 0,\\
B_{\alpha[\beta\gamma]} &= \tfrac{1}{2}c_{\alpha\beta\gamma} + \pi_\alpha{}^{\bar{\alpha}}(y)\,c_{\bar{\alpha}\beta\gamma}.
\end{aligned} \tag{4.4}$$

In these expressions the following notations are used:

$$\begin{aligned}
E_{[a}{}^m E_{b]}{}^n &= E_a{}^m E_b{}^n - E_b{}^m E_a{}^n,\\
F_{mn}{}^{\hat{\alpha}} &= \partial_m A_n{}^{\hat{\alpha}} - \partial_n A_m{}^{\hat{\alpha}} - A_m{}^{\hat{\beta}} A_n{}^{\hat{\gamma}} c_{\hat{\gamma}\hat{\beta}}{}^{\hat{\alpha}},\\
\pi_\alpha{}^{\bar{\alpha}}(y) &= e_\alpha{}^\mu(y)\,e_\mu{}^{\bar{\alpha}}(y),
\end{aligned} \tag{4.5}$$

and the structure constants of G are defined by

$$[Q_{\hat{\alpha}}, Q_{\hat{\beta}}] = c_{\hat{\alpha}\hat{\beta}}{}^{\hat{\gamma}} Q_{\hat{\gamma}}.$$

On substituting the expressions (4.4) into the $4+K$-dimensional curvature scalar one obtains

$$R_{4+K} = R_4 - \tfrac{1}{4}F_{ab}{}^{\hat{\alpha}} F_{ab}{}^{\hat{\beta}} D_{\hat{\alpha}}{}^{\gamma}(L_y)\,D_{\hat{\beta}}{}^{\gamma}(L_y) + R_K, \tag{4.6}$$

where $R_4 = R_4(E)$ is the usual 4-dimensional curvature scalar, expressed in terms of the vierbein $E_m{}^a(x)$, and R_K is the constant curvature of G/H (see Appendix II). On integrating y^μ the Yang–Mills Lagrangian emerges, since

$$\frac{1}{V_K}\int d^K y \det e(y)\,D_{\hat{\alpha}}{}^{\gamma}(L_y)\,D_{\hat{\beta}}{}^{\gamma}(L_y) = k\delta_{\hat{\alpha}\hat{\beta}},$$

where $k = \dim(G/H)/\dim(G)$. Thus, the above assertion that the system contains the graviton and massless Yang–Mills fields is verified.

There may also be zero mass scalars in the system, like the Brans–Dicke scalar in the $K=1$ theory. This point is sensitive, however, to the matter field couplings which are generally needed to force the compactification. We shall not therefore discuss the scalar modes.

Finally, to illustrate how the ansatz described above can be combined with an analogous one for a spinor field to give an effective 4-dimensional term, we consider the expression

$$\mathscr{L}_\psi = \frac{i}{2} \det E(x, y)\, \bar{\psi} \gamma^A E_A{}^M \nabla_M \psi + \text{h.c.} \tag{4.7}$$

with $\psi(x, y)$ represented by the ansatz,

$$\psi(x, y) = D(L_y{}^{-1})\, \psi(x), \tag{4.8}$$

This is a typical term from the harmonic expansion of $\psi(x, y)$. The spinor of $SO(1, 3+K)$ must of course be decomposed relative to the subgroup $SO(1, 3) \times H$; each piece in the decomposition will yield a $\psi(x)$ and the details of labelling are ignored here. One must substitute the expressions (4.1), (4.4) and (4.8) into (4.7),

$$\begin{aligned}\mathscr{L}_\psi = \frac{i}{2} \det E(x) \det e(y)\, \bar{\psi}(x)\, D(L_y) \Big[\gamma^a \Big\{ E_a{}^m \partial_m + E_a{}^\mu \partial_\mu - \frac{1}{4} B_{a[BC]} \gamma^{BC} \Big\} \\ + \gamma^\alpha \Big\{ e_\alpha{}^\mu \partial_\mu - \frac{1}{4} B_{\alpha[BC]} \gamma^{BC} \Big\} \Big] D(L_y{}^{-1})\, \psi(x) + \text{h.c.}\end{aligned} \tag{4.9}$$

To simplify this expression it is necessary to eliminate the derivative ∂_μ by means of the formula

$$\begin{aligned}\partial_\mu D(L_y{}^{-1}) &= -D(L_y{}^{-1})\, \partial_\mu D(L_y)\, D(L_y{}^{-1}) \\ &= -e_\mu{}^{\hat{\alpha}}(y)\, D(Q_{\hat{\alpha}})\, D(L_y{}^{-1})\end{aligned}$$

which follows from the definition (3.3). Multiplication by $e_\alpha{}^\mu$ then gives

$$e_\alpha{}^\mu \partial_\mu D(L_y{}^{-1}) = -(D(Q_\alpha) + \pi_\alpha{}^{\bar{\alpha}} D(Q_{\bar{\alpha}}))\, D(L_y{}^{-1}),$$

where π is defined in (4.5). It turns out that the π-containing terms are cancelled by similar terms in the spin connections. This cancellation is ensured by the embedding of H in the tangent space group (see Appendix IV). After some tedious calculations it is found that (4.9) reduces to

$$\begin{aligned}\mathscr{L}_\psi = \frac{i}{2} \det E(x) \det e(y)\, \bar{\psi}(x)\, D(L_y) \Big[\gamma^a \Big\{ E_a{}^m(x)\, \partial_m + \frac{1}{4} B_{a[bc]}(x)\, \gamma^{bc} \\ - A_a{}^{\hat{\alpha}}(x)\, D(Q_{\hat{\alpha}}) \Big\} + \frac{3}{8} F_{ab}{}^{\S}(x)\, \gamma^{ab} \gamma^\alpha D_\alpha{}^{\S}(L_y{}^{-1}) \\ - \gamma^\alpha D(Q_\alpha) + \frac{1}{8} c_{\alpha\beta\gamma} \gamma^{\alpha\beta\gamma} \Big] D(L_y{}^{-1})\, \psi(x) + \text{h.c.}\end{aligned} \tag{4.10}$$

and it remains only to integrate over y. Since γ^a must commute with $D(L_y)$, the result takes the form

$$\frac{i}{2} V_K \det E(x)\, \bar{\psi}(x) \left[\gamma^a \left\{E_a{}^m(x)\, \partial_m + \frac{1}{4} B_{a[bc]}(x)\, \gamma^{bc} - A_a{}^{\hat{\beta}}(x)\, D(Q_{\hat{\beta}})\right\}\right.$$

$$\left. + \frac{3}{8} F_{ab}{}^{\hat{\delta}}(x)\, \gamma^{ab} \Gamma^{\hat{\delta}} - M\right] \psi(x) + \text{h.c.}, \tag{4.11}$$

where $\Gamma^{\hat{\delta}}$ and M are matrices defined by

$$\Gamma^{\hat{\delta}} = \frac{1}{V_K} \int_{G/H} d\mu D(L_y)\, \gamma^\alpha D(L_y{}^{-1})\, D_\alpha{}^{\hat{\delta}}(L_y{}^{-1}), \tag{4.12}$$

$$M = \frac{1}{V_K} \int_{G/H} d\mu D(L_y)(\gamma^\alpha D(Q_\alpha) - \frac{1}{8} c_{\alpha\beta\gamma} \gamma^{\alpha\beta\gamma})\, D(L_y{}^{-1}). \tag{4.13}$$

Note that the "Pauli-moment term" implied by non-zero $\Gamma^{\hat{\delta}}$ is a necessary consequence of our formalism, and the precise numerical expressions for $\Gamma^{\hat{\delta}}$ and M are the hall-marks of the expansion on G/H.

V. CONCLUSIONS

In this study we have attempted to analyze in some detail various formal aspects of Kaluza–Klein theories in which the ground state geometry takes the form $M^4 \times B^K$, where B^K is a K-dimensional quotient space G/H. The symmetries of this ground state are used to classify the excitations of the system, and these excitations are found to include particles of spins zero, one and two, in general. In particular, there are a set of four vectors belonging to the adjoint representation of G contained among the components of the metric tensor and spin connection which serve as Yang–Mills potentials.

Harmonic expansions on B^K are used to define sequences of component fields, defined on M^4 and belonging to irreducible representations of the Yang–Mills group, G. An important feature in determining the content of these expansions is the embedding of the stability group H in the tangent space group $SO(K)$. This embedding is discussed in detail. Considerable use is made of the "boosts," L_y, in setting up the formalism. The transformation properties of L_y are used to define the action of G on the manifold G/H, and the derivatives of L_y give the covariant basis vectors on this manifold. They also are used in the construction of integral formulae for mass matrices and coupling parameters for Pauli-moment-like terms in the 4-dimensional effective theory (formulae (4.12) and (4.13)).

We remark that the ansatz for the 4-dimensional fields associated with massless particles—graviton, Yang–Mills vectors and Brans–Dicke-type scalars—can be seen to arise as leading terms in the harmonic expansions. The complete spectrum for the

$4+1$-dimensional theory is discussed in Appendix V. It is shown there that the massive excitations are purely of spin-2 and that they can be assigned to infinite-dimensional representations of the non-compact group $SO(1, 2)$. Such non-compact symmetries are spontaneously broken and may be viewed as spectrum generating symmetries. It is interesting that, viewed in this light, Kaluza–Klein theory seems to constitute an example of a consistent theory of massive, charged spin-2 particles.

Information drawn from the structure of the harmonic expansions would be useful in testing the classical stability of Kaluza–Klein type solutions. We have not dealt with this type of investigation (except for the simple example of $4+1$ theory) since it is generally necessary to introduce matter fields of some sort into the system to generate non-vanishing curvature in G/H.[5] Such matter fields, which must affect the stability considerations, represent a non-geometrical element in the theory (except, perhaps, in the case of extended supergravitities [7]) and are particular to specific models.

APPENDIX I: Non-linear Realizations

To supplement the discussion of Section III we give here a somewhat more amplified description of the properties of non-linear realizations of continuous groups. The concepts were thoroughly exposed in the physics literature a few years ago when their appropriateness in treating spontaneously broken symmetries was noticed [12]. In that work the space on which the group acts is a set of scalar fields and the non-linearity can be expressed through algebraic constraints on these fields. Because of such constraints it is impossible for all components to vanish simultaneously, hence the absence of an invariant ground state, and hence the relevance to spontaneous symmetry breaking. From a more abstract point of view, the set of scalar fields constitute a vector space which supports a linear representation of the symmetry. The constraints serve to pick out a subspace, generally curved, which is carried into itself by the action of the group. On such a subspace the group is said to be realized non-linearly. Indeed, to characterize the fields belonging to such a curved subspace it is generally necessary to adopt a curvilinear co-ordinate system and so to express the group actions by the transformations which they induce in these co-ordinates.

The classic example of non-linear realizations is the treatment of chiral $SU(2) \times SU(2)$. This group has the infinitesimal structure of $SO(4)$ and the scalar fields $(\sigma, \boldsymbol{\pi})$ are assumed to transform as a 4-vector. If they are subjected to the constraint

$$\sigma^2 + \boldsymbol{\pi}^2 = F^2, \tag{I.1}$$

where F is a fixed number, then the number of independent components is reduced to three: they belong to a 3-sphere of radius F. This 3-sphere is of course carried into itself by the transformations of $SO(4)$. It is invariant. To parametrize the points of the 3-sphere one could simply use the 3-vector $\boldsymbol{\pi}$ and write $\sigma = \pm\sqrt{F^2 - \boldsymbol{\pi}^2}$. For some purposes, however, it is more useful to define a set of boosts, L_π, which are $SO(4)$ transformations that carry a given reference point, say

---

[5] The original Kaluza–Klein theory in $4+1$-dimensions represents a case where G/H is flat. Generalizations of this have been investigated by Scherk and Schwarz [13].

$$(\sigma, \boldsymbol{\pi}) = (F, \mathbf{0}) \tag{I.2}$$

into a general point. In terms of the 4×4 orthogonal matrices $D_{\alpha\beta}(g)$ of the vector representation one writes

$$\begin{aligned} \sigma &= D_{44}(L_\pi)\, F, \\ \pi_i &= D_{i4}(L_\pi)\, F. \end{aligned} \tag{I.3}$$

Of course there is a great deal of latitude in the choice of L_π and it will generally be necessary to use at least two co-ordinate patches if the sphere is to be covered without generating artificial singularities.

To illustrate the parametrization of the sphere we give the example of spherical polar co-ordinates,

$$\begin{aligned} \sigma &= F\cos\theta, \\ \pi_1 &= F\sin\theta\cos\varphi, \\ \pi_2 &= F\sin\theta\sin\varphi\cos\psi, \\ \pi_3 &= F\sin\theta\sin\varphi\sin\psi, \end{aligned} \tag{I.4}$$

which corresponds to the choice of L_π,

$$L_\pi = e^{-\psi Q_{23}} e^{-\varphi Q_{12}} e^{-\theta Q_{41}}. \tag{I.5}$$

Here the operator $Q_{\alpha\beta} = -Q_{\beta\alpha}$ is the generator of the infinitesimal rotations in the $\alpha\beta$-plane

$$[Q_{\alpha\beta}, Q_{\gamma\delta}] = \delta_{\beta\gamma} Q_{\alpha\delta} - \delta_{\alpha\gamma} Q_{\beta\delta} + \delta_{\alpha\delta} Q_{\beta\gamma} - \delta_{\beta\delta} Q_{\alpha\gamma}. \tag{I.6}$$

On applying an $SO(4)$ transformation, g, to the sphere, the point whose co-ordinates are (θ, φ, ψ) will be carried into one with co-ordinates $(\theta', \varphi', \psi')$. A rotation through the angle ω in the 14-plane, for example, gives

$$\begin{pmatrix} \cos\theta' \\ \sin\theta'\cos\varphi' \end{pmatrix} = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta\cos\varphi \end{pmatrix}.$$

By solving these equations one can discover the transformed boost, $L_{\pi'}$. In general the co-ordinate transformation is given by

$$\begin{aligned} D_{\alpha 4}(L_{\pi'}) &= D_{\alpha\beta}(g)\, D_{\beta 4}(L_\pi) \\ &= D_{\alpha 4}(gL_\pi). \end{aligned}$$

From this it follows that the precise relation between L_π and $L_{\pi'}$ must take the form

$$gL_\pi = L_{\pi'} h, \tag{I.7}$$

where h is an orthogonal transformation in the 123-subspace, i.e.,

$$D_{\alpha 4}(h) = \delta_{\alpha 4}. \tag{I.8}$$

Since $L_{\pi'}$ is known once the angles $(\theta', \varphi', \psi')$ have been computed, the $SO(3)$ transformation, $h = L_{\pi'}{}^{-1} g L_\pi$, is well defined.

To compute the basis vectors on the 3-sphere consider the differential

$$\begin{aligned} e &= L_\pi{}^{-1} dL_\pi \\ &= e^{\theta Q_{41}} e^{\varphi Q_{12}} e^{\psi Q_{23}} d(e^{-\psi Q_{23}} e^{-\varphi Q_{12}} e^{-\theta Q_{41}}) \\ &= -d\theta Q_{41} - d\varphi(Q_{12}\cos\theta + Q_{42}\sin\theta) \\ &\quad - d\psi(Q_{23}\cos\varphi + (Q_{13}\cos\theta - Q_{43}\sin\theta)\sin\varphi) \\ &= -d\theta Q_{41} - d\varphi\sin\theta Q_{42} + d\psi\sin\theta\sin\varphi Q_{43} \\ &\quad - d\varphi Q_{12} - d\psi\cos\varphi Q_{23} + d\psi\cos\theta\sin\varphi Q_{31}. \end{aligned} \tag{I.9}$$

From this expression one extracts the coefficients

$$e_\mu{}^\alpha = \begin{array}{c} \\ \theta \\ \varphi \\ \psi \end{array} \begin{array}{c} \begin{array}{ccc} 41 & 42 & 43 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 \\ 0 & -\sin\theta & 0 \\ 0 & 0 & \sin\theta\sin\varphi \end{pmatrix} \end{array},$$

$$e_\mu{}^{\bar\alpha} = \begin{array}{c} \\ \theta \\ \varphi \\ \psi \end{array} \begin{array}{c} \begin{array}{ccc} 12 & 23 & 31 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -\cos\varphi & \cos\theta\sin\varphi \end{pmatrix} \end{array}, \tag{I.10}$$

where the labelling of rows and columns is indicated. The metric tensor on the 3-sphere is given by

$$\begin{aligned} ds^2 &= (d\theta e_\theta{}^\alpha + d\varphi e_\varphi{}^\alpha + d\psi e_\psi{}^\alpha)^2 \\ &= d\theta^2 + \sin^2\theta\, d\varphi^2 + \sin^2\theta\sin^2\varphi\, d\psi^2 \end{aligned} \tag{I.11}$$

as would be expected in spherical polar co-ordinates. The spin connection $B_{\alpha[\beta\gamma]}$ is constructed out of the quantities

$$\pi_\alpha{}^{\bar\beta} = e_\alpha{}^\mu e_\mu{}^{\bar\beta}$$

$$= \begin{array}{c} \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccc} 12 \quad & \quad 23 \quad & \quad 31 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ \dfrac{1}{\sin\theta} & 0 & 0 \\ 0 & -\dfrac{\cot\varphi}{\sin\theta} & \cot\theta \end{pmatrix} \end{array} \tag{I.12}$$

as discussed in the text, (4.4) (and further in Appendix 2).

The case discussed above is an example of a non-linear realization in which the group $G = SO(4)$ is made to act on a 3-sphere. The points of the 3-sphere are put into correspondence with the left cosets $SO(4)/SO(3)$ through the parametrization of the boosts, L_π. The procedure generalizes directly to manifolds of the form G/H, where H is some subgroup of G. (Right cosets can also be used. One writes

$$g\colon L_y \to L_{y'} = hL_y g^{-1}.$$

On the manifold of G itself one can define the action of $G \times G$

$$(g_1, g_2)\colon L_y \to L_{y'} = g_1 L_y g_2{}^{-1}.$$

In all such cases the mapping $y \to y'$ is determined once the form of L_y has been specified.)

Global properties of a quotient space G/H are often made transparent by embedding in a flat space of higher dimension, i.e., by viewing the non-linear realization as obtained from a linear one with constraints applied. A simple illustration of this is provided by the case of $SU(3)/SU(2) \times U(1)$ which we consider briefly.

The infinitesimal transformations of $SU(3)$ are generated by the eight operators

$$Q_a{}^b = -Q_b{}^{a\dagger}, \qquad a, b = 1, 2, 3 \qquad (Q_a{}^a = 0). \tag{I.13}$$

The boosts may be given by

$$L_y = \exp(y^\alpha Q_\alpha{}^3 - \text{h.c.}), \tag{I.14}$$

where the co-ordinates y^α, $\alpha = 1, 2$, are complex. The little group in this case, $H = SU(2) \times U(1)$, is generated by $Q_\alpha{}^\beta$. With this choice of L_y the procedures discussed above could be used to find the mapping $y^\alpha \to y'^\alpha$ and the induced $SU(2) \times U(1)$ transformation, h, corresponding to an $SU(3)$ transformation, g. However, we shall not pursue this.

The manifold $SU(3)/SU(2) \times U(1)$ clearly has two complex dimensions. In order to see that it is, in fact, $\mathbb{C}P_2$ one can proceed as follows. Let the hermitian, traceless matrices, $Y_a{}^b$, be the co-ordinates of a flat space on which $SU(3)$ acts linearly

$$Y \to Y' = g\,Y g^{-1}. \tag{I.15}$$

An invariant subspace can be picked out by requiring the co-ordinates to satisfy the matrix equations

$$Y^2 = \frac{R}{3} Y + \frac{2}{9} R^2, \tag{I.16}$$

where R is a constant. Since Y is hermitian, it can be diagonalized by an $SU(3)$ transformation and one finds easily that its diagonalized form must be

$$\hat{Y} = \begin{pmatrix} -R/3 & & \\ & -R/3 & \\ & & 2R/3 \end{pmatrix}. \tag{I.17}$$

In other words, the general solution of the constraints is given by

$$Y = L_y \hat{Y} L_y{}^{-1} \tag{I.18}$$

with appropriate $SU(3)$ boosts, L_y. On the other hand, it is possible to express the general solution in the form

$$Y_a{}^b = \left(\frac{1}{3} \delta_a{}^b - \frac{\eta_a \bar{\eta}^b}{\bar{\eta}\eta} \right) R, \tag{I.19}$$

where η_a is a complex triplet and $\bar{\eta}^a = \eta_a^*$. Since Y is unchanged by the complex scaling $\eta_a \to \lambda \eta_a$, this means that the manifold parametrized by L_y can be identified with the 2-dimensional complex projective space, $\mathbb{C}P_2$.

As a supplementary remark we note that it is possible to define the action of a non-compact group on a compact manifold. A trivial example of this is provided by the group of displacements on the real line. By factoring out a discrete subgroup corresponding to displacements which are integer multiples of some unit, $2\pi R$, one obtains a compact quotient space. Formally, one writes

$$L_y = e^{yQ}, \qquad 0 \leqslant y \leqslant 2\pi R, \tag{I.20}$$

where Q is the generator of infinitesimal displacements. Applying an arbitrary displacement, $g = \exp \omega Q$, on the left, one obtains as usual,

$$gL_y = L_{y'} h,$$

where

$$\begin{aligned} y' &= y + \omega - 2\pi R n, \\ h &= (e^{2\pi R Q})^n, \end{aligned} \tag{I.21}$$

and the integer n is determined such that y' lies in the interval $(0, 2\pi R)$. If the points $y=0$ and $y=2\pi R$ are identified then the quotient space is a circle of radius R.

A less trivial example is the realization of the non-compact $SO(1,2)$ by its action on the projective cone, a compact 1-dimensional manifold. That such a realization should exist is easily seen by embedding the cone in 3-dimensional Minkowski space. The null cone

$$Y_0{}^2 - Y_1{}^2 - Y_2{}^2 = 0 \tag{I.22}$$

is clearly invariant under the action of $SO(1,2)$. But it is a 2-dimensional space. However, the cone is also invariant under the scaling transformation

$$Y_a \to \lambda Y_a. \tag{I.23}$$

A 1-dimensional manifold is obtained by identifying those points which are related by scaling. (In practice one would be dealing with functions on the cone which are homogeneous of degree zero.) A possible parametrization of the manifold would involve writing

$$Y_a = \begin{pmatrix} 1 \\ \sin y \\ \cos y \end{pmatrix}. \tag{I.24}$$

To find the mapping $y \to y'$ corresponding to an hyperbolic rotation in the 02-plane, for example, one would write

$$Y'_a = \begin{pmatrix} ch\omega + sh\omega \cos y \\ \sin y \\ sh\omega + ch\omega \cos y \end{pmatrix} = \begin{pmatrix} \lambda \\ \lambda \sin y' \\ \lambda \cos y' \end{pmatrix}$$

to obtain the relation

$$\sin y' = \frac{\sin y}{ch\omega + sh\omega \cos y}. \tag{I.25}$$

This kind of realization can be formalized in the manner used previously. The generators, $Q_{ab} = -Q_{ba}$, of $SO(1,2)$ satisfy the commutation rules

$$[Q_{ab}, Q_{cd}] = \eta_{bc} Q_{ad} - \eta_{ac} Q_{bd} + \eta_{ad} Q_{bc} - \eta_{bd} Q_{ac}. \tag{I.26}$$

It is possible to pick out two of these,

$$Q_{02} \quad \text{and} \quad Q_+ = Q_{01} + Q_{12} \tag{I.27}$$

which are to be identified as the generators of the (non-compact) stability group, H. They satisfy

$$[Q_{02}, Q_+] = Q_+. \tag{I.28}$$

The boosts are defined by

$$L_y = e^{yQ_{12}}. \tag{I.29}$$

Since any element of $SO(1, 2)$ can be represented unambiguously in the form

$$e^{\alpha Q_{12}} e^{\beta Q_+} e^{\gamma Q_{02}} \tag{I.30}$$

it will always be possible to solve the mapping relation

$$gL_y = L_{y'}h$$

for y' and h as functions of y and g.

Another example is the flat group of Scherk and Schwarz [13]. In its simplest version, the generator Q_1 is distinguished from the others, Q_j, $j = 2, 3, \ldots, K$. The commutation rules are

$$\begin{aligned} [Q_i, Q_j] &= 0, \\ [Q_i, Q_1] &= M_{ij} Q_j, \end{aligned} \tag{I.31}$$

where M is antisymmetric, i.e., Q_1 generates orthogonal transformations while the Q_i generate translations. (For $K = 3$ this is simply the 2-dimensional Poincaré algebra.) To construct a compact manifold for the group to act on, consider the elements

$$L_y = e^{y^1 Q_1} e^{y^j Q_j}, \tag{I.32}$$

where y^1 ranges from 0 to $2\pi/m_1$ ($m_1 =$ smallest non-zero eigenvalue of iM_{jk}) and the y^j are restricted to the ranges

$$0 \leqslant y^j \leqslant 2\pi R^j, \qquad j = 2, \ldots, N. \tag{I.33}$$

The action of the group on L_y is easily found. Firstly, the action of $\exp \omega^1 Q_1$ is given by

$$e^{\omega^1 Q_1} L_y = L_{y'}, \tag{I.34}$$

where $y'^j = y^j$ and $y'^1 = y^1 + \omega^1$ (mod $2\pi/m_1$). Secondly,

$$e^{\omega^j Q_j} L_y = L_{y'}h, \tag{I.35}$$

where $y'^1 = y^1$ and

$$y'^j = y^j + (e^{-y^1 M})_{jk}\, \omega^k - 2\pi R^j n^j \tag{I.36}$$

with the integers n^j chosen such that y'^j falls in the interval $(0, 2\pi R^j)$. The little group transformation associated with this is then

$$h = \prod_j (e^{2\pi R^j Q_j})^{n_j}. \tag{I.37}$$

The covariant basis is defined as usual by the 1-form

$$\begin{aligned} e &= L_y{}^{-1} dL_y \\ &= dy^1(Q_1 - y^j M_{jk} Q_k) + dy^j Q_j, \end{aligned}$$

i.e.,

$$e_\mu{}^\alpha = \begin{array}{c|cc} & Q_1 & Q_k \\ \hline y^1 & 1 & -y^i M_{ik} \\ y^j & 0 & \delta_{jk} \end{array}. \tag{I.38}$$

To construct Killing vectors the matrices $D_\alpha{}^\beta$ of the adjoint representation are needed. These are given by

$$\begin{aligned} D_1{}^\alpha(L_y)\, Q_\alpha &= L_y{}^{-1} Q_1 L_y \\ &= Q_1 - y^i M_{ik} Q_k, \\ D_j{}^\alpha(L_y)\, Q_\alpha &= L_y{}^{-1} Q_j L_y \\ &= (e^{y^1 M})_{jk}\, Q_k, \end{aligned}$$

i.e.,

$$D_\alpha{}^\beta(L_y) = \begin{pmatrix} 1 & -y^i M_{ik} \\ 0 & (\exp y^1 M)_{jk} \end{pmatrix}. \tag{I.39}$$

The Killing vectors for left translations are then

$$\begin{aligned} K_\alpha{}^\mu(y) &= D_\alpha{}^\beta(L_y)\, e_\beta{}^\mu \\ &= \begin{pmatrix} 1 & 0 \\ 0 & (\exp y^1 M)_{jk} \end{pmatrix}. \end{aligned} \tag{I.40}$$

(This matrix was denoted $U(y)$ in the work of Scherk and Schwarz.)

We close with the observation that the zero-torsion spin connection $B_\alpha{}^\beta$ (see Appendix 2) is given for this manifold by

$$B_1{}^j = -B_j{}^1 = 0, \qquad B_j{}^k = -dy^1 M_{jk}, \tag{I.41}$$

and the curvature 2-form vanishes, i.e., the manifold is indeed flat.

APPENDIX II: CURVATURE AND TORSION ON G/H

Because of their high degree of symmetry, quotient spaces have rather simple geometrical properties and it is possible to derive a number of explicit formulae to describe these. In this appendix we give derivations for such of these formulae as have been referred to in the text. All of them arise out of manipulations of the boost elements L_y which we use to characterize the manifold G/H.

As has been mentioned in Section III, the first and most important step is to construct the vectors of a covariant basis on the manifold. To this end one must consider the 1-form

$$\begin{aligned} e(y) &= L_y{}^{-1} dL_y \\ &= e^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}} \\ &= dy^{\mu} e_{\mu}{}^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}}, \end{aligned} \tag{II.1}$$

where the infinitesimal generators of G are denoted $Q_{\hat{\alpha}}$. These generators satisfy the commutation rules

$$[Q_{\hat{\alpha}}, Q_{\hat{\beta}}] = c_{\hat{\alpha}\hat{\beta}}{}^{\hat{\gamma}} Q_{\hat{\gamma}} \tag{II.2}$$

with structure constants $c^{\hat{\alpha}\hat{\beta}\hat{\gamma}}$.

The differential properties of $e(y)$ needed for the discussion of curvature and torsion are summed up in the 2-form

$$\begin{aligned} de(y) &= -dL_y{}^{-1} \wedge dL_y \\ &= L_y{}^{-1} dL_y L_y{}^{-1} \wedge dL_y \\ &= e(y) \wedge e(y), \end{aligned} \tag{II.3}$$

the Cartan–Maurer formula. On substituting the expression $e = e^{\hat{\alpha}} Q_{\hat{\alpha}}$ and using the commutation rules satisfied by the generators, one can extract the equivalent form

$$de^{\hat{\alpha}} = \tfrac{1}{2} e^{\hat{\beta}} \wedge e^{\hat{\gamma}} c_{\hat{\beta}\hat{\gamma}}{}^{\hat{\alpha}} \tag{II.4}$$

or, in terms of components,

$$\partial_{\mu} e_{\nu}{}^{\hat{\alpha}} - \partial_{\nu} e_{\mu}{}^{\hat{\alpha}} = -e_{\mu}{}^{\hat{\beta}} e_{\nu}{}^{\hat{\gamma}} c_{\hat{\beta}\hat{\gamma}}{}^{\hat{\alpha}}. \tag{II.5}$$

As explained in the text the generators $Q_{\hat{\alpha}}$ separate naturally into two sets, $Q_{\bar{\alpha}}$ belonging to the subgroup H, and Q_{α}, the remainder. The coefficients $e_{\mu}{}^{\alpha}$, $\alpha = 1, 2, \ldots, N$, constitute a set of N covariant basis vectors. The reciprocal basis is denoted $e_{\alpha}{}^{\mu}$, i.e.,

$$e_{\alpha}{}^{\mu} e_{\mu}{}^{\beta} = \delta_{\alpha}{}^{\beta}, \qquad e_{\mu}{}^{\alpha} e_{\alpha}{}^{\nu} = \delta_{\mu}{}^{\nu}. \tag{II.6}$$

The torsion 2-form, T^α, is defined by

$$T^\alpha = de^\alpha + e^\beta \wedge B_\beta{}^\alpha, \tag{II.7}$$

where the 1-form $B_\beta{}^\alpha$ represents the spin connection. Once the latter quantity has been assigned it is possible to construct, in addition to T^α, the curvature 2-form $R_\alpha{}^\beta$ defined by

$$R_\alpha{}^\beta = dB_\alpha{}^\beta + B_\alpha{}^\gamma \wedge B_\gamma{}^\beta. \tag{II.8}$$

One of the simplest possibilities is to choose B such that the torsion vanishes,

$$0 = de^\alpha + e^\beta \wedge B_\beta{}^\alpha. \tag{II.9}$$

On comparing this equation with the Cartan–Maurer formula and noting that $c_{\bar\beta\bar\gamma}{}^\alpha = 0$ (because H is a group) one finds immediately

$$\begin{aligned} B_\beta{}^\alpha &= e^\gamma \tfrac{1}{2} c_{\gamma\beta}{}^\alpha + e^{\bar\gamma} c_{\bar\gamma\beta}{}^\alpha \\ &= e^\gamma(\tfrac{1}{2} c_{\gamma\beta}{}^\alpha + \pi_\gamma{}^{\bar\gamma} c_{\bar\gamma\beta}{}^\alpha). \end{aligned} \tag{II.10}$$

In this last equation we have used

$$e^{\bar\gamma} = e^\gamma \pi_\gamma{}^{\bar\gamma} = e^\gamma e_\gamma{}^\mu e_\mu{}^{\bar\gamma}. \tag{II.11}$$

From the expression (II.10) for B one obtains the curvature 2-form

$$R_\beta{}^\gamma = \tfrac{1}{2} e^\alpha \wedge e^\epsilon (\tfrac{1}{2} c_{\alpha\epsilon}{}^\delta c_{\delta\beta}{}^\gamma + c_{\alpha\epsilon}{}^{\bar\delta} c_{\bar\delta\beta}{}^\gamma + \tfrac{1}{2} c_{\alpha\beta}{}^\delta c_{\epsilon\delta}{}^\gamma). \tag{II.12}$$

(In deriving this we have used $c_{\alpha\bar\beta}{}^{\bar\gamma} = 0$ which is not generally true. See for example, the discussion of $SO(2, 1)$ in Appendix I. In the general case one would have to keep such terms.)

Another simple possibility is to choose B such that the torsion takes the form

$$T^\alpha = \tfrac{1}{2} e^\beta \wedge e^\gamma c_{\beta\gamma}{}^\alpha. \tag{II.13}$$

Comparison with the Cartan–Maurer formula this time gives

$$B_\beta{}^\alpha = e^{\bar\gamma} c_{\bar\gamma\beta}{}^\alpha \tag{II.14}$$

and hence the curvature

$$R_\beta{}^\gamma = \tfrac{1}{2} e^\alpha \wedge e^\epsilon c_{\alpha\epsilon}{}^{\bar\delta} c_{\bar\delta\beta}{}^\gamma. \tag{II.15}$$

In particular, if the manifold is G itself, i.e., $H = 1$, then this assignment gives vanishing curvature. (On the other hand, for some interesting manifolds such as $SO(N + 1)/SO(N)$, the structure constants $c_{\alpha\beta}{}^\gamma$ vanish and the two assignments are identical.)

APPENDIX III: TRANSFORMATION PROPERTIES

Some symmetry properties of the manifold G/H are needed for an understanding of the transformation behaviour of the Yang–Mills fields. The purpose of this appendix is to discuss the derivation of the necessary formula.

Consider firstly the transformation rule (3.8) for the 1-form $e^{\hat\alpha}$. For an x-dependent left translation, $y \to y'(x, y)$, this reads

$$e^{\hat\alpha}(y') = e^{\hat\beta}(y)\, D_{\hat\beta}{}^{\hat\alpha}(h^{-1})^{\hat\alpha} + (g^{-1}dg)^{\hat\beta}\, D_{\hat\beta}{}^{\hat\alpha}(L_y h^{-1}), \tag{III.1}$$

where $D_{\hat\alpha}{}^{\hat\beta}$ is a matrix of the adjoint representation, i.e.,

$$g^{-1}Q_{\hat\alpha}\, g = D_{\hat\alpha}{}^{\hat\beta}(g)\, Q_{\hat\beta}.$$

The components $(g^{-1}dg)^{\hat\alpha}$ and $(hdh^{-1})^{\hat\alpha}$ are defined by

$$\begin{aligned} g^{-1}dg &= (g^{-1}dg)^{\hat\alpha}\, Q_{\hat\alpha}, \\ hdh^{-1} &= (hdh^{-1})^{\hat\alpha}\, Q_{\hat\alpha} \\ &= (hdh^{-1})^{\bar\alpha}\, Q_{\bar\alpha}, \end{aligned} \tag{III.2}$$

which is possible since $g^{-1}dg$ belongs to the algebra of G while hdh^{-1} belongs to the subalgebra associated with H. In the following we shall assume that the algebra is fully reducible, i.e.,

$$D_{\alpha}{}^{\bar\beta}(h) = D_{\bar\alpha}{}^{\beta}(h) = 0. \tag{III.3}$$

This means that (III.1) separates into components from G/H and H, respectively,

$$e^{\alpha}(y') = e^{\beta}(y)\, D_{\beta}{}^{\alpha}(h^{-1}) + (g^{-1}dg)^{\hat\beta}\, D_{\beta}{}^{\alpha}(L_y h^{-1}), \tag{III.4}$$

$$e^{\bar\alpha}(y') = e^{\bar\beta}(y)\, D_{\bar\beta}{}^{\bar\alpha}(h^{-1}) + (hdh^{-1})^{\bar\alpha} + (g^{-1}dg)^{\hat\beta}\, D_{\hat\beta}{}^{\bar\alpha}(L_y h^{-1}). \tag{III.5}$$

Under the transformation $y \to y'(x, y)$ we have

$$dy'^{\mu} = dy^{\nu}\, \frac{\partial y'^{\mu}}{\partial y^{\nu}} + dx^{m}\, \frac{\partial y'^{\mu}}{\partial x^{m}}$$

or, equivalently,

$$dy^{\mu} = dy'^{\nu}\, \frac{\partial y^{\mu}}{\partial y'^{\nu}} + dx^{m}\, \frac{\partial y^{\mu}}{\partial x^{m}}.$$

Partial derivatives transform such that

$$\begin{aligned} dy^{\mu}\,\partial_{\mu} + dx^{m}\partial_{m} &= dy'^{\mu}\left(\frac{\partial y^{\nu}}{\partial y'^{\mu}}\,\partial_{\nu}\right) + dx^{m}\left(\partial_{m} + \frac{\partial y^{\nu}}{\partial x^{m}}\,\partial_{\nu}\right) \\ &= dy'^{\mu}\,\partial'_{\mu} + dx^{m}\,\partial'_{m} \end{aligned}$$

(where ∂'_{m} denotes differentiation with respect to x^{m} at fixed y'^{μ}).

Extracting the coefficients of dy'^{μ} and dx^m from the formulae (III.4) and (III.5) gives

$$e_\mu{}^\alpha(y') = \frac{\partial y^\nu}{\partial y'^\mu}\, e_\nu{}^\beta(y)\, D_\beta{}^\alpha(h^{-1}), \tag{III.6}$$

$$0 = \frac{\partial y^\nu}{\partial x^m}\, e_\nu{}^\beta(y)\, D_\beta{}^\alpha(h^{-1}) + (g^{-1}\partial_m g)^{\hat\beta}\, D_{\hat\beta}{}^\alpha(L_y h^{-1}), \tag{III.7}$$

$$e_\mu{}^{\bar\alpha}(y') = \frac{\partial y^\nu}{\partial y'^\mu}\, (e_\nu{}^{\bar\beta}(y)\, D_{\bar\beta}{}^{\bar\alpha}(h^{-1}) + (h\partial_\nu h^{-1})^{\bar\alpha}), \tag{III.8}$$

$$\begin{aligned} 0 = {} & \frac{\partial y^\nu}{\partial x^m}\, e_\nu{}^{\bar\beta}(y)\, D_{\bar\beta}{}^{\bar\alpha}(h^{-1}) + (h\partial'_m h^{-1})^{\bar\alpha} \\ & + (g^{-1}\partial_m g)^{\hat\beta}\, D_{\hat\beta}{}^{\bar\alpha}(L_y h^{-1}). \end{aligned} \tag{III.9}$$

The formula (III.7) can be arranged in the form

$$\frac{\partial y^\mu}{\partial x^m} = -(g^{-1}\partial_m g)^{\hat\beta}\, K_{\hat\beta}{}^\mu(y), \tag{III.10}$$

where $K_{\hat\beta}{}^\mu$, the so-called Killing vector, is defined by the expression

$$K_{\hat\beta}{}^\mu(y) = D_{\hat\beta}{}^\gamma(L_y)\, e_\gamma{}^\mu(y). \tag{III.11}$$

Its transformation behaviour is obtained with the help of (III.6), viz.

$$\begin{aligned} K_{\hat\beta}{}^\mu(y') &= D_{\hat\beta}{}^\gamma(L_{y'})\, e_\gamma{}^\mu(y') \\ &= D_{\hat\beta}{}^\gamma(gL_y h^{-1})\, D_\gamma{}^\alpha(h)\, e_\alpha{}^\nu(y)\, \frac{\partial y'^\mu}{\partial y^\nu} \\ &= D_{\hat\beta}{}^{\hat\gamma}(g)\, K_{\hat\gamma}{}^\nu((y)\, \frac{\partial y'^\mu}{\partial y^\nu}. \end{aligned} \tag{III.12}$$

From (III.10) and (III.12) it is a simple exercise to derive an expression for $\partial y'^\mu/\partial x^m$. Thus,

$$\begin{aligned} \frac{\partial y'^\mu}{\partial x^m} &= -\frac{\partial y^\nu}{\partial x^m}\frac{\partial y'^\mu}{\partial y^\nu} \\ &= (g^{-1}\partial_m g)^{\hat\beta}\, K_{\hat\beta}{}^\nu(y)\, \frac{\partial y'^\mu}{\partial y^\nu} \\ &= (g^{-1}\partial_m g)^{\hat\beta}\, D_{\hat\beta}{}^{\hat\gamma}(g^{-1})\, K_{\hat\gamma}{}^\mu(y') \\ &= -(g\partial_m g^{-1})^{\hat\beta}\, K_{\hat\beta}{}^\mu(y'). \end{aligned} \tag{III.13}$$

With the help of (III.12) and (III.13) one can easily extract the transformation rule for the Yang–Mills gauge fields $A_a{}^{\hat\beta}(x)$ defined by (4.1), or

$$E_a{}^\mu(x,y) = A_a{}^{\hat\beta}(x)\,K_{\hat\beta}{}^\mu(y). \tag{III.14}$$

Since $E_a{}^\mu$ is part of a contravariant vector it must transform according to

$$E_a{}^\mu(x,y) \rightarrow E_a'^{\mu}(x,y') = E_a{}^\nu(x,y)\frac{\partial y'^\mu}{\partial y^\nu} + E_a{}^n(x,y)\frac{\partial y'^\mu}{\partial x^n}. \tag{III.15}$$

Substitution of (III.14) into this rule gives

$$\begin{aligned} A_a'^{\hat\beta}(x)\,K_{\hat\beta}{}^\mu(y') &= A_a{}^{\hat\beta}(x)\,K_{\hat\beta}{}^\nu(y)\frac{\partial y'^\mu}{\partial y^\nu} + E_a{}^n(x)\frac{\partial y'^\mu}{\partial x^n} \\ &= A_a{}^{\hat\beta}(x)\,D_{\hat\beta}{}^{\hat\gamma}(g^{-1})\,K_{\hat\gamma}{}^\mu(y') - E_a{}^n(x)(g\partial_n g^{-1})^{\hat\beta}\,K_{\hat\beta}{}^\mu(y'). \end{aligned}$$

Since $K_{\hat\beta}{}^\mu(y')$ is a common factor in this equation, it may be removed. One is left with the rule

$$A_a'^{\hat\beta}(x) = A_a{}^{\hat\gamma}(x)\,D_{\hat\gamma}{}^{\hat\beta}(g^{-1}) - E_a{}^n(x)(g\partial_n g^{-1})^{\hat\beta}$$

or, equivalently,

$$A_m'(x) = gA_m(x)\,g^{-1} - g\partial_m g^{-1}, \tag{III.16}$$

where $A_m(x) = E_m{}^a(x)\,A_a{}^{\hat\beta}(x)\,Q_{\hat\beta}$.

A more direct construction of the Killing vectors makes reference to infinitesimal transformations. Thus, the infinitesimal form of the transformation rule $L_{y'} = gL_y h^{-1}$ reads

$$L_{y+\delta y} = (1 + \delta g^{\hat\alpha} Q_{\hat\alpha})\,L_y(1 - \delta h^{\bar\alpha} Q_{\bar\alpha}),$$

i.e.,

$$\delta y^\mu\,\partial_\mu L_y = \delta g^{\hat\alpha} Q_{\hat\alpha} L_y - L_y \delta h^{\bar\alpha} Q_{\bar\alpha}$$

to first order in the infinitesimals δy^μ, $\delta g^{\hat\alpha}$ and $\delta h^{\bar\alpha}$. Multiply on the left by $L_y{}^{-1}$ and use the definition of e,

$$\begin{aligned} \delta y^\mu e_\mu{}^{\hat\alpha} Q_{\hat\alpha} &= \delta g^{\hat\alpha} L_y{}^{-1} Q_{\hat\alpha} L_y - \delta h^{\bar\alpha} Q_{\bar\alpha} \\ &= \delta g^{\hat\alpha} D_{\hat\alpha}{}^{\hat\beta}(L_y)\,Q_{\hat\beta} - \delta h^{\bar\alpha} Q_{\bar\alpha}. \end{aligned}$$

Extract the coefficient of Q_α

$$\delta y^\mu e_\mu{}^\alpha = \delta g^{\hat\beta} D_{\hat\beta}{}^\alpha(L_y)$$

since $\delta h^{\alpha}=0$. Hence the infinitesimal δy^{μ} is given by

$$\begin{aligned}\delta y^{\mu} &= \delta g^{\beta} D_{\beta}{}^{\alpha}(L_y)\, e_{\alpha}{}^{\mu}(y)\\ &= \delta g^{\beta} K_{\beta}{}^{\mu}(y)\end{aligned} \tag{III.17}$$

thereby defining the Killing vectors associated with left translations on G/H.

The same procedure could be followed for constructing the Killing vectors associated with right translations. One should use the 1-form $\tilde{e}=LdL^{-1}$ instead of e to obtain $\tilde{K}_{\hat{\alpha}}{}^{\mu}=D_{\hat{\alpha}}{}^{\beta}(L_y{}^{-1})\,\tilde{e}_{\beta}{}^{\mu}$. For the special case, $H=1$, the result is particularly simple. One finds $\tilde{K}_{\hat{\alpha}}{}^{\mu}=-e_{\hat{\alpha}}{}^{\mu}$.

APPENDIX IV: HARMONIC EXPANSIONS

The general form of harmonic expansions on the manifold G/H was discussed in Section III. In particular, for the functions $\psi_i(x,y)$ which transform under the combined left translations, $y\rightarrow y'$, and associated tangent space rotations, h, such that

$$\psi_i(x,y)\rightarrow \psi'_i(x,y')=\mathbb{D}_{ij}(h)\,\psi_j(x,y) \tag{IV.1}$$

the appropriate expansion would be

$$\psi_i(x,y)=\sum_n \sqrt{\frac{d_n}{d_{\mathbb{D}}}}\sum_{\zeta,p} D^n_{i\zeta,p}((L_y{}^{-1})\,\psi^n_{p\zeta}(x). \tag{IV.2}$$

The sum includes all irreducible representations, D^n, of G which contain $\mathbb{D}_{ij}$ on restriction to the subgroup, H. If $\mathbb{D}_{ij}$ occurs more than once in D^n, then the supplementary label ζ is needed. The matrices $D_{qp}^{\,n}$ have dimension d_n. The expansion coefficients ψ^n are fields on 4-dimensional spacetime which belong to the irreducible representation D^n,

$$\psi^n_{p\zeta}(x)\rightarrow \psi'^{\,n}_{p\zeta}(x)=D_{pq}^{\,n}(g)\,\psi_{q\zeta}^{\,n}(x). \tag{IV.3}$$

They can be projected from $\psi_i(x,y)$ by integrating over the manifold

$$\psi_{p\zeta}^{\,n}(x)=\frac{1}{V_K}\sqrt{\frac{d_n}{d_{\mathbb{D}}}}\int_{G/H} d\mu D^n_{p,i\zeta}(L_y)\,\psi_i(x,y). \tag{IV.4}$$

Differentiation with respect to y^{μ} of the fields $\psi_i(x,y)$ is equivalent to an algebraic operation on the components $\psi^n(x)$. The simplest covariant derivative that incorporates $\partial/\partial y^{\mu}$ would be

$$\begin{aligned}\nabla_{\alpha}\psi_i(x,y) &= e_{\alpha}{}^{\mu}\,\partial_{\mu}\psi_i-\tfrac{1}{2}B_{\alpha[\beta\gamma]}\mathbb{D}_{ij}(\Sigma^{\beta\gamma})\,\psi_j\\ &= e_{\alpha}{}^{\mu}\,\partial_{\mu}\psi_i+e_{\alpha}{}^{\mu}e_{\mu}{}^{\bar{\beta}}\mathbb{D}_{ij}(Q_{\bar{\beta}})\,\psi_j.\end{aligned} \tag{IV.5}$$

In this formula we have used the expression

$$B_{\alpha[\beta\gamma]} = e_\alpha{}^\mu e_\mu{}^{\bar{\alpha}} c_{\bar{\alpha}\beta\gamma},$$

which is one of the possible spin connections on G/H, derived in Appendix II.[6] We have also used the formula (3.17) which fixes the embedding of H in the tangent space group, $SO(N)$,

$$-\tfrac{1}{2} c_{\bar{\alpha}\beta\gamma} \mathbb{D}(\Sigma^{\beta\gamma}) = \mathbb{D}(Q_{\bar{\alpha}}).$$

Differentiation of the boost $L_y{}^{-1}$ leads back to the definitions (3.3) of e, viz.

$$\begin{aligned} \partial_\mu L_y{}^{-1} &= -L_y{}^{-1} \partial_\mu L_y L_y{}^{-1} \\ &= -e_\mu{}^\beta Q_\beta L_y{}^{-1}, \end{aligned}$$

and this implies that, in any representation,

$$\partial_\mu D^n(L^{-1}) = -e_\mu{}^\beta D^n(Q_\beta)\, D^n(L_y{}^{-1}). \tag{IV.6}$$

On comparing this with (IV.5) one finds for the covariant derivative of D^n,

$$\nabla_\alpha D^n_{ip}(L^{-1}) = -D^n_{iq}(Q_\alpha)\, D^n_{qp}(L^{-1}). \tag{IV.7}$$

This formula was used in Section IV to reduce some terms from the $4+N$-dimensional Lagrangian to 4-dimensional form.

Finally, the expansion formulae are sketched for two relativity simple cases.

(1) $SU(2)/U(1)$. This is the case of spherical harmonics on the 2-dimensional sphere. With $G = SU(2)$ generated by $Q_{ab} = -Q_{ba}$, $a, b = 1, 2, 3$ and $H = U(1)$ generated by Q_{12} we can take $y^\mu = (\theta, \varphi)$ with

$$L_y = e^{-\varphi Q_{12}} e^{-\theta Q_{31}}. \tag{IV.8}$$

The covariant basis vectors are contained in the 1-form

$$\begin{aligned} e &= e^{\theta Q_{31}} e^{\varphi Q_{12}} d(e^{-\varphi Q_{12}} e^{-\theta Q_{31}}) \\ &= -d\varphi \cos\theta Q_{12} + d\varphi \sin\theta Q_{23} - d\theta Q_{31}. \end{aligned} \tag{IV.9}$$

Expressed in matrix notation, with rows and columns indicated,

$$e_\mu{}^{\hat{\alpha}} = \begin{array}{c} \\ \theta \\ \varphi \end{array} \begin{array}{c} \begin{matrix} 13 & 23 & 12 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin\theta & -\cos\theta \end{pmatrix} \end{array}. \tag{IV.10}$$

The last column gives the components of $e_\mu{}^{12}(= e_\mu{}^{\bar{\alpha}}$ in the notation of Section III).

---

[6] This spin connection corresponds to non-vanishing torsion on G/H. In general the spin connection will contain terms additional to this one, including fluctuation terms, for example.

The tangent space group, $O(2)$, in this case coincides with $H = U(1)$ so the embedding problem is trivial. The 2-vector ϕ_α resolves into helicity $\lambda = \pm 1$ combinations,

$$\frac{1}{\sqrt{2}}(\phi_1 \pm i\phi_2),$$

and likewise for higher rank tensors. For a function, $\phi_\lambda(x, y)$ of helicity λ, integer or half-integer, with respect to the tangent space group we have the expansion

$$\phi_\lambda(x, y) = \sum_j \sqrt{2j+1} \sum_m D^{\,j}_{\lambda m}(L_y{}^{\ 1})\, \phi^{\,j}_{m\lambda}(x), \tag{IV.11}$$

where j takes the values $|\lambda|$, $|\lambda| + 1, \ldots$ and D^j denotes the $2j+1$-dimensional representation of $SU(2)$,

$$\begin{aligned} D^{\,j}_{\lambda m}(L_y{}^{-1}) &= D^{\,j}_{\lambda m}(e^{\theta Q_{31}} e^{\varphi Q_{12}}) \\ &= d^{\,j}_{\lambda m}(\theta)\, e^{im\varphi}, \end{aligned} \tag{IV.12}$$

in the notation of Wigner.

Generalization to the N-dimensional sphere $SO(N+1)/SO(N)$ is straightforward in principle though, of course, complicated in detail. There is no embedding to be done since $H = SO(N)$ coincides with the tangent space group. Our second example illustrates a case where this is not so.

(2) $SU(3)/SU(2) \times U(1)$. With $G = SU(3)$ generated by the octet of charges, $Q_a{}^b = -Q_b{}^{a\dagger}$ ($Q_a{}^a = 0$), $a, b = 1, 2, 3$ and H generated by the subset $Q_\alpha{}^\beta$, $\alpha, \beta = 1, 2$ we can parametrize G/H by a complex 2-vector

$$L_y = e^{y^1 Q_1{}^3 + y^2 Q_2{}^3 - \mathrm{h.c.}} \tag{IV.13}$$

This would not be the most convenient choice for practical computations but we do not intend to present more than a sketch here. The question of interest is how to embed $H = SU(2) \times U(1)$ in the tangent space group, $SO(4)$. The general formula for this, (3.17), could be applied but it is simpler in the present case to work *ab initio*.

The triplet of $SU(3)$ decomposes under the subgroup $SU(2) \times U(1)$ such that

$$3 = 2_{1/3} + 1_{-2/3}, \tag{IV.14}$$

i.e., into an $SU(2)$ doublet with $U(1)$ quantum number 1/3 and a singlet with $-2/3$. For the octet the decomposition reads

$$8 = 1_0 + 3_0 + 2_1 + 2_{-1}. \tag{IV.15}$$

If the octet is real then the doublets 2_1 and 2_{-1} are related by complex conjugation. In particular, the generators $Q_\alpha{}^3$, $Q_3{}^\alpha$ associated with G/H are in the representations 2_1 and 2_{-1}, respectively. According to the general principle expressed in formula

(3.17), the 4-vector of $SO(4) \sim SU(2) \times SU(2)$ must have the $SU(2) \times U(1)$ content of G/H,

$$(2, 2) = 2_1 - 2_{-1}. \tag{IV.16}$$

This means that the embedding must be such that

$$\begin{aligned} (2, 1) &= 2_0, \\ (1, 2) &= 1_1 + 1_{-1}, \end{aligned} \tag{IV.17}$$

for the 2-spinors of $SO(4)$. An explicit construction for the $SU(2) \times U(1)$ decomposition of a 4-vector, ϕ_α, is easily arranged. From the 2×2 matrix

$$\begin{pmatrix} \phi_4 + i\phi_3 & i(\phi_1 - i\phi_2) \\ i(\phi_1 + i\phi_2) & \phi_4 - i\phi_3 \end{pmatrix} \tag{IV.18}$$

select the first column. This belongs to the representation 2_1, and so we define the spinor ϕ_λ with components

$$\begin{aligned} \phi_{1/2} &= \frac{1}{\sqrt{2}} (\phi_4 + i\phi_3), \\ \phi_{-1/2} &= \frac{1}{\sqrt{2}} (i\phi_1 - \phi_2). \end{aligned} \tag{IV.19}$$

Now the representation 2_1 is found in the $SU(3)$ multiplets 8, 10, 27,..., with triality zero. It follows, therefore, that the harmonic expansion of the 4-vector, $\phi_\alpha(x, y)$, must take the form

$$\phi_\lambda(x, y) = \sum_{n = 8, 10, 27, \ldots} \sqrt{\frac{d_n}{2}} \sum_p D^n_{\lambda p}(L_y{}^{-1})\, \phi_p{}^n(x), \tag{IV.20}$$

where the coefficients $\phi_p{}^n(x)$ are complex. Expansions for tensors of $SO(4)$ can be constructed by straightforward generalization of this procedure. It is interesting to note, however, that spinors of $SO(4)$ cannot be represented in this way. This is clear from the fact that the $SU(2) \times U(1)$ content of the 2-spinors, viz. 2_0, 1_1 and $1_{\ 1}$ is not to be found in any $SU(3)$ multiplet. In fact, it is not possible to define spinors on the manifold $SU(3)/SU(2) \times U(1)$.

APPENDIX V: SPECTRUM OF 4 + 1 KALUZA–KLEIN THEORY

Since the ground state of the $K = 1$ theory is flat, it can be obtained as a vacuum solution of the 4 + 1-dimensional Einstein equations without a cosmological term. For $K > 1$ the manifold G/H has constant curvature and it is necessary to introduce non-geometrical (matter) field to act as sources. Because of the simplicity of the

$K=1$ theory one can easily obtain its excitation spectrum and relate it to an underlying non-compact symmetry, $SO(1,2)$.

From the fact that the $4+1$-dimensional Einstein theory has five degrees of freedom one expects that all massive states belong to multiplets of spin 2. This is easily verified. Write the metric tensor in the form

$$g_{MN}=\eta_{MN}+h_{MN}, \tag{V.1}$$

where $M, N=0, 1, 2, 3, 4$ and η_{MN} is the flat metric,

$$\eta_{MN}=\operatorname{diag}(+1,-1,-1,-1,-1). \tag{V.2}$$

Treating the components h_{MN} as small quantities, one obtains the second order terms in the Einstein Lagrangian

$$\begin{aligned}\mathscr{L}_2=&\tfrac{1}{4}(h_{MN,L}h_{MN,L}-2h_{MN,L}h_{ML,N}+2h_{MN,N}h_{LL,M}-h_{MM,L}h_{NN,L})\\&+\tfrac{1}{2}J_{MN}h_{MN},\end{aligned} \tag{V.3}$$

where J_{MN} is an external source. The connections vanish so that simple partial derivatives are indicated, $h_{MN,L}=\partial_L h_{MN}$. The equations of motion derived from (V.3) take the form

$$h_{MN,LL}-h_{ML,NL}-h_{NL,ML}+h_{LL,MN}+\eta_{MN}h_{KL,KL}-\eta_{MN}h_{KK,LL}=J_{MN}. \tag{V.4}$$

These equations are compatible only if J is conserved,

$$J_{MN,N}=0. \tag{V.5}$$

This is a relic of the 5-dimensional general covariance. The usual way to solve (V.4), subject to the compatibility condition (V.5), is to impose a co-ordinate condition. For example, in the gauge

$$h_{MN,N}=0 \tag{V.6}$$

they are solved by

$$h_{MN}=\frac{1}{\partial^2}\left[J_{MN}-\frac{1}{3}\left(\eta_{MN}-\frac{\partial_M\partial_N}{\partial^2}\right)J_{LL}\right], \tag{V.7}$$

where ∂^2 represents the 5-dimensional d'Alembertian.

On substituting the solution (V.7) back into the Lagrangian (V.3), discarding total derivatives, one finds that $\mathscr{L}_2$ reduces to

$$\begin{aligned}\mathscr{L}_2&=\frac{1}{4}J_{MN}h_{MN}\\&=\frac{1}{4}\left[J_{MN}\frac{1}{\partial^2}J_{MN}-\frac{1}{3}J_{MM}\frac{1}{\partial^2}J_{NN}\right].\end{aligned} \tag{V.8}$$

This expression represents the effective interaction between conserved (i.e., physical) sources due to the exchange of the Kaluza–Klein particles.

Now consider the pole, $\partial^2 = 0$, in (V.8). The residue simplifies on using the conservation of J_{MN}. Firstly, if $\partial_4 \neq 0$ we can choose a frame in which

$$\partial_0 = \partial_4 \qquad \text{and} \qquad \partial_i = 0, \qquad i = 1, 2, 3. \tag{V.9}$$

In this frame we have

$$J_{0M} = J_{4M}, \qquad M = 0, 1, 2, 3, 4 \tag{V.10}$$

so that $J_{MM} = -J_{ii}$ and $J_{MN}^2 = J_{ij}^2$. It follows that, in the neighbourhood of the pole, the effective interaction (V.8) reduces to

$$\mathscr{L}_2 \simeq \frac{1}{4} J_{ij}^t \frac{1}{\partial_0^2 - \partial_4^2} J_{ij}^t, \tag{V.11}$$

where the physically significant part of the source is the traceless 3-dimensional tensor

$$J_{ij}^t = J_{ij} - \tfrac{1}{3}\delta_{ij} J_{kk}. \tag{V.12}$$

This result indicates that the massive states ($\partial_4 \neq 0$) are indeed purely spin 2. Their masses are given by the eigenvalues of $-\partial_4^2 = (n/2\pi R)^2$, $n = 1, 2, \ldots$, where R is the radius of the Kaluza–Klein circle.

On the other hand, if $\partial_4 = 0$ we have the massless states and they can be analysed conveniently by choosing a frame where

$$\partial_0 = \partial_3 \qquad \text{and} \qquad \partial_1 = \partial_2 = \partial_4 = 0. \tag{V.13}$$

In this frame the effective interaction reduces, in the neighbourhood of the pole, to the form

$$\frac{1}{4} \sum_{\lambda=-2}^{2} J_{-\lambda} \frac{1}{\partial_0^2 - \partial_3^2} J_{\lambda}, \tag{V.14}$$

where λ refers to the $O(2)$ helicity and

$$\begin{aligned}
J_2 &= \frac{1}{2}(J_{11} - J_{22}) + iJ_{12}, \\
J_1 &= J_{41} + iJ_{42}, \\
J_0 &= \frac{1}{\sqrt{6}}(J_{11} + J_{22} - 2J_{44}), \\
J_{-1} &= J_{41} - iJ_{42}, \\
J_{-2} &= \frac{1}{2}(J_{11} - J_{22}) - iJ_{12}.
\end{aligned} \tag{V.15}$$

The terms in (V.14) correspond to the exchange of graviton, photon and Bands–Dicke scalar.

Turning now to the non-compact symmetry aspect of this theory, we note that the general co-ordinate transformations in 4 + 1-dimensions include, in particular, the set

$$\begin{aligned}\delta x^m &= 0,\\ \delta y &= \omega^1(x)\frac{\cos My}{M} + \omega^2(x)\frac{\sin My}{M} + \omega^3(x)\frac{1}{M},\end{aligned} \tag{V.16}$$

where $M = 1/2\pi R$. (The finite transformations are discussed in Appendix I. See, in particular (I.25).) The infinitesimal generators are given by

$$Q_1 = -\frac{\cos My}{M}\partial_y, \qquad Q_2 = -\frac{\sin My}{M}\partial_y, \qquad Q_3 = -\frac{1}{M}\partial_y, \tag{V.17}$$

and these are easily seen to generate the algebra of $SO(1, 2)$,

$$[Q_1, Q_2] = -Q_3, \qquad [Q_2, Q_3] = Q_1, \qquad [Q_3, Q_1] = Q_2. \tag{V.18}$$

Moreover, the Casimir invariant vanishes,

$$\begin{aligned}\eta^{ab}Q_aQ_b &= -\left(\frac{\cos My}{M}\partial_y\right)^2 - \left(\frac{\sin My}{M}\partial_y\right)^2 + \left(\frac{1}{M}\partial_y\right)\\ &= 0.\end{aligned}$$

This means that fields defined on the circle belong to an irreducible representation of $SO(1, 2)$. It is infinite dimensional but not unitary, however. Since the mass operator is given by $-\partial_y{}^2 = M^2Q_3^2$ one sees that the spectrum of masses is simply the spectrum of Q_3^2 in the irreducible representation of $SO(1, 2)$.

Generalizations to higher dimensional manifolds are possible. For example, on the sphere S^K, which is invariant under $SO(K + 1)$, it is possible to define the action of $SO(1, K + 1)$. However, to find the excitation spectrum and perhaps relate it to the spectrum of some combination of generators of $SO(1, K + 1)$, it would be necessary to include the excitations of the matter fields wich are needed for the compactification.

ACKNOWLEDGMENT

The authors wish to thank S. Randjbar-Daemi and W. Mecklenburg for many stimulating and helpful discussions.

References

1. Th. Kaluza, *Sitz. Preuss. Akad. Wiss. Berlin, Math. Phys. K*1 (1921), 966.
2. O. Klein, *Z. Phys.* **37** (1926), 895.
3. J. Scherk and J. H. Schwarz, *Phys. Lett. B* **57** (1975), 463; E. Cremmer and J. Scherk, *Nucl. Phys. B* **103** (1976), 393; **108** (1976), 409.
4. B. De Witt, "Dynamical Theory of Groups and Fields," Gordon & Breach, New York, 1965.
5. Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, *Nucl. Phys. B* **127** (1977), 57; L. Palla, "Proc. 1978 Tokyo Conf. on High Energy Physics," p. 629.
6. Z. Horvath and L. Palla, *Nucl. Phys. B* **142** (1978), 327.
7. P. G. O. Freund and M. A. Rubin, *Phys. Lett. B* **97** (1980), 233.
8. A. S. Schwarz, *Comm. Math. Phys.* **56** (1977), 79; V. N. Romanov, A. S. Schwarz, and Yu. S. Tyupkin, *Nucl. Phys. B* **130** (1977), 209; V. N. Romanov, I. V. Frolov, and A. S. Schwarz, *Teor. Mat. Fiz.* **37** (1978), 305; A. S. Schwarz and Yu. S. Tyupkin, *Nucl. Phys. B* **187** (1981), 321; R. Casalbuoni, G. Domokos, and S. Kovesi–Domokos, *Phys. Rev. D* **17** (1978), 2048.
9. J. F. Luciani, *Nucl. Phys. B* **135** (1978), 111; E. Witten, *Nucl. Phys. B* **186** (1981), 412.
10. R. Kerner, *Ann. Inst. H. Poincaré* **9** (1968), 143; Y. M. Cho and P. G. O. Freund, *Phys. Rev. D* **12** (1975), 1711; Y. M. Cho, *J. Math. Phys.* **16** (1975), 2029; Y. M. Cho and P. S. Jang, *Phys. Rev. D* **12** (1975), 2789; A. Trautman, *Rep. Math. Phys.* **16** (1975), 2029; K. I. Macrae, *Phys. Rev. D* **18** (1978), 3737; C. Orzalesi, Multidimensional unified theories, Parma preprint IFPR/TH 63 (1980); W. Mecklenburg, *Phys. Rev. D* **21** (1980), 2149.
11. K. Nomizu, *Amer. J. Math.* **76** (1954), 33; S. Kobayashi and K. Nomizu, "Foundations of Differential Geometry," Vol. II, Wiley–Interscience, New York, 1969.
12. See, for example, Abdus Salam and J. Strathdee, *Phys. Rev.* **184** (1969), 1750; **184** (1969), 1760.
13. J. Scherk and J. H. Schwarz, *Nucl. Phys. B* **153** (1979), 61.

CHAPTER

4

Supergravity

SUPERGRAVITY THEORY IN 11 DIMENSIONS

E. CREMMER, B. JULIA and J. SCHERK
Laboratoire de Physique Théorique de l'Ecole Normale Supérieure [1], *Paris, France*

Received 4 April 1978

We present the action and transformation laws of supergravity in 11 dimensions which is expected to be closely related to the O(8) theory in 4 dimensions after dimensional reduction.

Extended O(N) (N = 1, ...8) supergravity theories [1–5] are notoriously difficult to construct beyond $N = 3$. The difficulty lies partly in the large number of fields involved (for $N = 8$, which is the largest theory that can be constructed in this frame-work, one has 1 graviton, 8 spin 3/2 gravitinos, 28 vectors, 56 spinors, 35 scalar and 35 pseudoscalar particles) but mostly in the fact that the spin 0 fields appear in a non-polynomial way, thus forbidding a step-by-step construction of the action and transformation laws. So far, only the $N = 4$ theory has been constructed in a closed form [3], the simplest form of it exhibiting a manifest SU(4) invariance [5], while the $N = 8$ theory [4] has been constructed only to order K^2. Further, geometrical methods [6] do not seem readily applicable to these theories beyond $N = 2$, due to the presence of fields not bearing a vector index, which makes their interpretation as gauge fields difficult.

On the other hand, an elegant method has been found to be very useful to circumvent the similar problem of constructing in a simple way the O(N) (N = 1, ...4) supersymmetric Yang–Mills theories in 4 dimensions. One first establishes the existence of a supersymmetric Yang–Mills theory in 10 dimensions [7], which was suggested by the study of the dual spinor model. Then one reduces the theory to four dimensions by assuming all fields to be independent of the extra 6 spacial coordinates. In this way, the $N = 4$ theory, which has a vanishing β-function at the first two non-trivial orders [8] was found, and also a systematic search of all supersymmetric Yang–Mills in less than 10 dimensions was conducted [9].

As shown by Nahm [10], $D = 10$ is the highest number of dimensions in which supersymmetry representations with $J \leqslant 1$ can exist, while supergravity theories ($J \leqslant 2$) can exist up to $D = 11$. The interest in constructing the 11 dimensional theory lies in the fact that its reduction to four dimensions is automatically guaranteed to yield an O(7) invariant supergravity theory which has exactly the same field content as the O(8) theory, both theories being presumably equivalent just as the O(4) and SU(4) supergravity theories have been shown to be.

The field content of the $D = 11$ theory is remarkably simple. It consists of the vierbein V_μ^a, a Majorana spin 3/2 ψ_μ, and of a completely antisymmetric gauge tensor with 3 indices $A_{\mu\nu\rho}$. To arrive at this set of fields a simple argument is to count the number of physical states. In $D = 11$, the Dirac matrices are 32×32 and a Majorana spin 3/2 field ψ_μ represents $1/2 \cdot 32(D-3) = 128$ degrees of freedom. The vierbein field represents $\{(D-1)(D-2)\}/2 - 1 = 44$ degrees of freedom. The mismatch is $84 = \binom{9}{3}$, which is just the number of components of a transverse, antisymmetric gauge field with 3 indices in 11 dimensions. Transversality amounts to requiring that beside coordinate invariance, local Lorentz invariance and local supersymmetry, the action is also invariant under the Abelian gauge transformation:

$$\delta A_{\mu\nu\rho} = \partial_\mu \zeta_{\nu\rho} + \partial_\nu \zeta_{\rho\mu} + \partial_\rho \zeta_{\mu\nu} \,,$$

---

[1] Laboratoire Propre du C.N.R.S., associé à l'Ecole Normale Supérieure et à l'Université de Paris-Sud. Postal address: 24 rue Lhomond, 75231 Paris Cedex 05, France.

where $\zeta_{\mu\nu} = -\zeta_{\nu\mu}$.

This gauge invariance, together with the requirement of absence of terms with more than two derivatives, implies that the action is polynomial in the $A_{\mu\nu\rho}$ field. This is a considerable simplification compared to the non-polynomial reduced forms.

Another way to arrive at this simplest set of fields is to start from the spinor dual model of closed strings. This model is obtained by doubling the Neveu–Schwarz (Bose states, NS)–Ramond (Fermi states, R) model [11] and a priori contains four possible sectors: NS ⊗ NS, NS ⊗ R, R ⊗ NS and R ⊗ R. In ref. [7] it was shown that the first sector, at zero mass, contains a graviton V^a_μ an antisymmetric tensor $A_{\mu\nu}$, and a scalar ϕ, while the second sector contains Majorana–Weyl spin 3/2 ψ_μ and spin 1/2 χ fields forming an irreducible representation of $D = 10$ supergravity, and the two other sectors were neglected. However, if we include them we get in addition a new spin 3/2 ψ'_μ and a spin 1/2 χ' field of opposite helicities and a bispinor field representing 64 degrees of freedom, equivalent to an antisymmetric gauge field with 3 indices $A_{\mu\nu\rho}$ (56 degrees of freedom) and a gauge vector field A_ρ (8 degrees of freedom). This set of fields forms an irreducible representation of extended supergravity in $D = 10$ dimensions, a point which was also realized by Schwarz [12]. It is easy to see how this set of fields arises from reducing the $D = 11$ theory to $D = 10$. The V^A_M decomposes into V^a_μ, A_ρ, ϕ fields (taking into account the fact that only the symmetric part propagates). The A_{MNP} decomposes into $A_{\mu\nu\rho}, A_{\mu\nu}$. Finally the Majorana ψ_M field decomposes into the Majorana–Weyl fields $\psi_\mu, \psi'_\mu, \chi, \chi'$.

It is also interesting to show that reducing the theory down to 4 dimensions we obtain the same counting of fields as in the O(8) theory. The Majorana ψ_M field decomposes into 8 spin 3/2 fields ψ^i_μ ($i = 1, \ldots 8$), and 56 spin 1/2 fields ψ^i_a ($i = 1, \ldots 8; a = 1, \ldots 7$). The graviton field decomposes into 1 graviton $g_{\mu\nu}$, 7 vectors $g_{\mu a}$ and 28 scalars ($g_{ab} = g_{ba}$). The antisymmetric tensor decomposes into an $A_{\mu\nu\rho}$ field equivalent, in 4 dimensions, to an auxiliary scalar field carrying no degrees of freedom, 7 antisymmetric $A_{\mu\nu a}$ gauge fields equivalent in 4 dimensions to 7 scalar fields ϕ_a, 21 vector fields $A_{\mu ab}$, and 35 pseudoscalar fields A_{abc}. The total content of fields is thus the same as in the O(8) theory, although dimensional reduction will only make an O(7) invariance manifest.

Let us now present the action and transformation laws of the $D = 11$ supergravity theory. Our metric is $(+ - - \ldots -)$; Greek indices are world indices while Latin indices refer to the tangent space. The Γ^a matrices are in the Majorana representations and form a purely imaginary representation of the Clifford algebra in 11 dimensions. $\Gamma^{a_1 \cdots a_N}$ represents the product of $N\,\Gamma$ matrices completely antisymmetrized, i.e. for unequal indices

$$\Gamma^{a_1 \ldots a_N} = \Gamma^{a_1} \cdots \Gamma^{a_N} .$$

The lagrangian we find is the following:

$$\mathcal{L} = -\frac{V}{4K^2} R(\omega)$$

$$-\frac{\mathrm{i}V}{2} \bar{\psi}_\mu \Gamma^{\mu\nu\rho} D_\nu \left(\frac{\omega+\hat{\omega}}{2}\right) \psi_\rho - \frac{V}{48} F_{\mu\nu\rho\sigma} F^{\mu\nu\rho\sigma}$$

$$+ \frac{KV}{192} (\bar{\psi}_\mu \Gamma^{\mu\nu\alpha\beta\gamma\delta} \psi_\nu + 12\, \bar{\psi}^\alpha \Gamma^{\gamma\delta} \psi^\beta)(F_{\alpha\beta\gamma\delta} + \hat{F}_{\alpha\beta\gamma\delta})$$

$$+ \frac{2K}{(144)^2} \epsilon^{\alpha_1\alpha_2\alpha_3\alpha_4\beta_1\beta_2\beta_3\beta_4\mu\nu\rho} F_{\alpha_1\alpha_2\alpha_3\alpha_4} F_{\beta_1\beta_2\beta_3\beta_4} A_{\mu\nu\rho}$$

The quartic terms in this Lagrangian are absorbed in the supercovariant fields (i.e. fields which under supersymmetry transform without derivatives of ϵ) $\hat{\omega}$ and $\hat{F}$ which will be defined below.

$F_{\mu\nu\rho\sigma}$ is the field strength associated with the gauge field $A_{\mu\nu\rho}$:

$$F_{\mu\nu\rho\sigma} = 4\,\partial_{[\mu} A_{\nu\rho\sigma]} ,$$

where the brackets represent the antisymmetrized sum over all permutations, divided by their number. The covariant derivative of ψ_μ is given by:

$$D_\nu(\omega)\psi_\mu = \partial_\nu \psi_\mu + \tfrac{1}{4}\, \omega_{\nu ab} \Gamma^{ab} \psi_\mu ,$$

and the convention for R is the same as in ref. [13].

The Lorentz connection coefficients $\omega_{\mu ab}$ are given by:

$$\omega_{\mu ab} = \omega^0_{\mu ab}(V) + K_{\mu ab} ,$$

where the contorsion tensor is

$$K_{\mu ab} = (\mathrm{i}K^2/4)[-\bar{\psi}_\alpha \Gamma_{\mu ab}{}^{\alpha\beta} \psi_\beta$$

$$+ 2(\bar{\psi}_\mu \Gamma_b \psi_a - \bar{\psi}_\mu \Gamma_a \psi_b + \bar{\psi}_b \Gamma_\mu \psi_\alpha)] .$$

The torsion tensor is given by:

$$C_{\mu\nu}{}^{a} = K_{\nu\mu}{}^{a} - K_{\mu\nu}{}^{a}$$
$$= (\mathrm{i}K^2/2)\,[\bar{\psi}_\alpha \Gamma_{\mu\nu}{}^{a\alpha\beta}\psi_\beta - 2\bar{\psi}_\mu \Gamma^a \psi_\nu]\,.$$

The transformation laws are given by:

$$\delta V^a_\mu = -\mathrm{i}K\bar{\epsilon}\Gamma^a\psi_\mu\,,$$

$$\delta\psi_\mu = \frac{1}{K}D_\mu(\hat{\omega})\epsilon + \frac{\mathrm{i}}{144}(\Gamma^{\alpha\beta\gamma\delta}{}_\mu$$
$$-8\Gamma^{\beta\gamma\delta}\delta^\alpha_\mu)\epsilon\hat{F}_{\alpha\beta\gamma\delta} \equiv \frac{1}{K}\hat{D}_\mu\epsilon\,,$$

$$\delta A_{\mu\nu\rho} = \tfrac{3}{2}\bar{\epsilon}\Gamma_{[\mu\nu}\psi_{\rho]}\,,$$
$$\hat{F}_{\mu\nu\rho\sigma} = F_{\mu\nu\rho\sigma} - 3K\bar{\psi}_{[\mu}\Gamma_{\nu\rho}\psi_{\sigma]}\,.$$

In order to obtain $\omega_{\mu ab}$ from its equation of motion $\hat{\omega}_{\mu ab}$ should be put equal to:

$$\hat{\omega}_{\mu ab} = \omega_{\mu ab} + (\mathrm{i}K^2/4)\bar{\psi}_\alpha\Gamma_{\mu ab}{}^{\alpha\beta}\psi_\beta\,.$$

As one can easily verify, $\hat{\omega}_{\mu ab}$ is a supercovariant tensor.

To obtain this action and transformation laws, we have proceeded in the following deductive fashion: Firstly, in the 1.5 order formalism and taking the linear part in δV and $\delta\psi_\mu$, the terms of the form $\bar{\epsilon}\psi$ in δS, the variation of the action vanishes as for $D = 4$ supergravity. Then, as we have a kinetic term for $A_{\mu\nu\rho}$ (whose scale has been fixed by the conventional factor 1/48), to cancel the $\bar{\epsilon}\psi F^2$, we need a $\bar{\psi}X\psi F$ coupling and a $Z\epsilon F$ term in $\delta\psi$, where X and Z are unknown tensors made of Γ matrices.

To determine these terms, it is shortest to use the requirement that the equations of motion of ψ_μ must be supercovariant. That is defining $\hat{D}_\mu$ by $\delta\psi_\mu = (1/K)\hat{D}_\mu\epsilon$ and the supercovariant field strength by $\psi_{\nu\rho} = \hat{D}_\nu\psi_\rho - \hat{D}_\rho\psi_\nu$, the fermion field equation should read:

$$\Gamma^{\mu\nu\rho}\hat{D}_\nu\psi_\rho = 0\,.$$

Looking at the terms of the form $Z\psi F$ in this equation and comparing them with what is obtained from the $\bar{\psi}X\psi F$ terms in the action fixes the form of the X and Z tensors and relate their coefficients.

Then we consider the terms of the form $\bar{\epsilon}\psi F^2$, in the variation of the action. They all indeed disappear if the coefficient of Z is fixed properly, except for one term involving a product of $9\,\Gamma$ matrices which can be cancelled by adding to S the gauge invariant expression:

$$a\int \mathrm{d}x\,\epsilon^{\alpha\beta\gamma\delta\,\mu\nu\rho\sigma ijk}F_{\alpha\beta\gamma\delta}F_{\mu\nu\rho\sigma}A_{ijk}\,,$$

and only if we have: $\delta A_{\mu\nu\rho} = b\bar{\epsilon}\Gamma_{[\mu\nu}\psi_{\rho]}$.

This fixes the product ab. Then looking at the terms of the type $\bar{\epsilon}\partial\psi F$ and $\bar{\epsilon}\psi\partial F$ in the variation, we determine b so that all terms are fixed, up to trilinear terms in $\delta\psi$ and up to quartic terms in the fermion fields in the Lagrangian.

In order to fix these, we require that $\psi_{\nu\rho}$ be indeed supercovariant. This imposes that we replace F by $\hat{F}$ in $\delta\psi$, but also ω by $\hat{\omega}$ since ω is not supercovariant. Now the transformation laws are fully fixed and a crucial test

transformation laws are fully fixed and a crucial test

is to see whether the supersymmetry algebra closes at least on the Bose fields. This is indeed the case, confirming the need to replace F and ω by their supercovariant versions and the correctness of the transformation laws.

Finally the quartic terms in the action are fixed so as to reproduce the supercovariant fermion field equation. The fact that this is possible is a test of consistency and requires the following identity derived by a Pauli–Fierz transformation:

$$\tfrac{1}{8}\Gamma^{\mu\nu\alpha\beta\gamma\delta}\psi_\nu\bar{\psi}_\alpha\Gamma_{\beta\gamma} - \tfrac{1}{8}\Gamma_{\beta\gamma}\psi_\nu\bar{\psi}_\alpha\Gamma^{\mu\nu\alpha\beta\gamma\delta}$$
$$-\tfrac{1}{4}\Gamma^{\mu\nu\alpha\beta\delta}\psi_\nu\bar{\psi}_\alpha\Gamma_\beta + \tfrac{1}{4}\Gamma_\beta\psi_\nu\bar{\psi}_\alpha\Gamma^{\mu\nu\alpha\beta\delta}$$
$$-2g^{\beta[\alpha}\Gamma^{\delta\mu\nu]}\psi_\nu\bar{\psi}_\alpha\Gamma_\beta - 2\Gamma_\beta\psi_\nu\bar{\psi}_\alpha g^{\beta[\alpha}\Gamma^{\delta\mu\nu]}$$
$$+2g^{\beta[\alpha}\Gamma^{\delta\mu\nu]}\bar{\psi}_\alpha\Gamma_\beta\psi_\nu = 0\,.$$

As usual the supersymmetry algebra closes on shell. Explicitly we find:

$$[\delta_2,\delta_1]V^a_\mu = \partial_\mu\xi^\nu V^a_\nu + \xi^\nu\partial_\nu V^a_\mu$$
$$-\mathrm{i}\kappa\bar{\epsilon}'\Gamma^a\psi_\mu + \Omega^a{}_b V_\mu{}^b\,,$$

where

$$\xi^\nu = -\mathrm{i}\bar{\epsilon}_1\Gamma^\nu\epsilon_2\,,\qquad \epsilon' = -K\xi^\nu\psi_\nu\,,$$

and

$$\Omega^a{}_b = (K/72)\bar{\epsilon}_1(\Gamma^{ab\alpha\beta\gamma\delta}$$
$$-24V_b{}^\alpha V^{a\beta}\Gamma^{\gamma\delta})\epsilon_2\hat{F}_{\alpha\beta\gamma\delta} + \xi^\nu\hat{\omega}_\nu{}^a{}_b\,,$$

$$[\delta_2,\delta_1]A_{\mu\nu\rho} = 3(\partial_{[\mu}\xi^\sigma)A_{\sigma\nu\rho]} + \xi^\sigma\partial_\sigma A_{\mu\nu\rho}$$
$$+\tfrac{3}{2}\bar{\epsilon}'\Gamma_{[\mu\nu}\psi_{\rho]} + \partial_{[\mu}\Lambda_{\nu\rho]}\,,$$

where

$$\Lambda_{\nu\rho} = (3/2K)\bar{\epsilon}_1 \Gamma_{\nu\rho}\epsilon_2 - 3\xi^\sigma A_{\sigma\nu\rho} \,,$$

$$[\delta_2, \delta_1]\,\psi_\mu = \partial_\mu \xi^\nu \psi_\nu + \xi^\nu \partial_\nu \psi_\mu$$

$$+ (1/K)\hat{D}_\mu \epsilon' + \tfrac{1}{4}\Omega_{ab}\Gamma^{ab}\psi_\mu + \mathcal{R}_\mu \,,$$

where $\mathcal{R}_\mu$ is proportional to the fermion equation of motion.

To check the full invariance of the action, we have successively verified the vanishing of terms of type: $\epsilon\psi^3(F+\hat{F}/2)$ and $\epsilon\psi^2 D(\hat{\omega})\psi$ to minimize the number of terms of the type $\epsilon\psi^5$ which have been finally shown analytically to cancel.

In conclusion, two things remain to be done with our theory. First, we are studying the reduction to four dimensions and the connection with the O(8) theory and also the reduction to 10 dimensions to get the zero slope limit of the closed string dual model of ref. [7]. The second is to find geometrical interpretations analogous to the one obtained for $D = 4$ in the case of O(1) and O(2) supergravity theories. The natural candidate for the graded Lie algebra is OSp(32, 1) which contains an internal O(8) subalgebra. Works along these lines are presently in progress.

References

[1] D.Z. Freedman, P. van Nieuwenhuizen and S. Ferrara, Phys. Rev. D13 (1976) 3214;
D.Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D14 (1976) 912;
S. Deser and B. Zumino, Phys. Lett. 62B (1976) 335;
S. Ferrara and P. van Nieuwenhuizen, Phys. Rev. Lett. 37 (1976) 1669.

[2] D.Z. Freedman, Phys. Rev. Lett. 38 (1976) 105;
S. Ferrara, J. Scherk and B. Zumino, Phys. Lett. 66B (1977) 35 and Nucl. Phys. B121 (1977) 393.

[3] A. Das, Phys. Rev. D15 (1977) 2805;
E. Cremmer, J. Scherk and S. Ferrara, Phys. Lett. 68B (1977) 234;
E. Cremmer and J. Scherk, Nucl. Phys. B127 (1977) 259.

[4] B. deWit and D.Z. Freedman, Nucl. Phys. B130 (1977) 105.

[5] E. Cremmer, J. Scherk and S. Ferrara, LPTENS 77/19 preprint, to be published in Physics Letters B.

[6] S.W. MacDowell and F. Mansouri, Phys. Rev. Lett. 38 (1977) 739;
A. Chamseddine and P.C. West, Nucl. Phys. B129 (1977) 34;
P.K. Townsend and P. van Nieuwenhuizen, Phys. Lett. 67B (1977) 439;
Y. Ne'eman and T. Regge, Univ. of Texas preprint (Dec. 1977).

[7] F. Gliozzi, J. Scherk and D. Olive, Nucl. Phys. B122 (1977) 253.

[8] D.R T. Jones, Phys. Lett. 72B (1977) 199;
E.C. Poggio and H.N. Pendleton, Phys. Lett. 72B (1977) 200.

[9] L. Brink, J.H. Schwarz and J. Scherk, Nucl. Phys. B121 (1977) 77.

[10] W. Nahm, CERN TH 2341 preprint (1977).

[11] A. Neveu and J.H. Schwarz, Nucl. Phys. B31 (1971) 86, Phys. Rev. D4 (1971) 1109;
P. Ramond, Phys. Rev. D3 (1971) 2415.

[12] J.H. Schwarz, private communication. He also computed some of the couplings and transformation laws of the $D = 10$ supergravity theory.

[13] D.Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D14 (1976) 912.

THE $N = 8$ SUPERGRAVITY THEORY. I. THE LAGRANGIAN

E. CREMMER and B. JULIA
Laboratoire de Physique Théorique de l'Ecole Normale Supérieure [1]*, 75231 Paris Cedex 05, France*

Received 25 September 1978

The SO(8) supergravity action is constructed in closed form. A local SU(8) group as well as the exceptional group E_7 are invariances of the equations of motion and of a new first order lagrangian.

Introduction. Supersymmetry became popular in elementary particle physics after the construction of the dual spinor model where the fermionic extension of the two dimensional conformal algebra was used to decouple the negative norm states from the physical sector of the Hilbert space. But the main appeal of its generalization to four dimensional theories lies in its predictive power, even if it remains to be seen whether this quality will manifest itself at the phenomenological level, or turn into a drawback and force us to study more complicated schemes. To answer this question, it is necessary to construct supersymmetric Lagrangians explicitly and to study their properties as classical and quantum field theories. Like all other continuous symmetries, supersymmetry has a local version; it is realized in the so-called extended supergravity theories all of which generalize bosonic gravity theories and unify gravitational interactions with other interactions, mediated by lower spin fields, like colour, weak and electromagnetic forces. The largest of these models has an SO(8) internal symmetry if one restricts oneself to one spin two elementary field, and no derivatives higher than two in the lagrangian. It contains all SO(N) modesl $N < 8$ including Einstein's theory ($N = 0$); its classical spectrum is the following: 1 graviton, 8 Rarita–Schwinger spin 3/2 fields, 28 vectors, 56 Majorana spin 1/2 fields, and 70 scalars, all massless. A very strong constraint arises for $N = 8$ because there is no "matter multiplet" to couple to the supergravity Lagrangian, i.e. one is forced to a true unification of all particles. However, this theory has a serious defect from the point of view of conservative quantum field theorists, it contains a coupling constant K with dimension of the inverse of a mass, and hence is not obviously renormalizable (one of the reasons being the non-polynomiality of the action: KA is dimensionless for all scalar fields A and appears to all powers). It is thus rather encouraging that supersymmetry partly eliminates this difficulty. There are some indications that all pure extended supergravity theories are one loop and probably two loop finite; the symmetry not only forbids the appearance of unwanted counterterms but restricts the lagrangians so much that the smallest non-polynomial theory, the SO(4) supergravity, has been constructed in closed form. It also permits in principle a step by step construction of the SO(8) theory, but algebraic complexity prevented the authors of ref. [1] to go beyond the second order in K. One must use more powerful techniques to obtain the full lagrangian in order to investigate classical spontaneous symmetry breaking or the finiteness of higher order quantum corrections.

Here come the idea of dimensional reduction and surprisingly the (two-dimensional) dual spinor model again. This remarkable model is though to be supersymmetric (in the 10 dim. embedding space-time), various truncated versions have been proved to be supersymmetric in the zero-slope limit [2], and the complete theory has been constructed in the same approximation by assuming supersymmetry [3]: this is the 11 dimensional supergravity, and the largest containing at most spin 2 particles after reduction from 11 to 4 dimensions [4]. The dimensional reduction implies a second low-energy approximation, beyond the zero-slope limit (which neglects extended modes of the string), it ignores (in its simplest version) all states with non-vanishing momentum in the internal directions x^i ($4 \leqslant i \leqslant 10$, say), that is (high) 4-dimensional mass.

---

[1] Laboratoire propre du C.N.R.S. associé à l'Ecole Normale Supérieure et à l'Université de Paris Sud.

To put our results in perspective let us recall two properties of models suggested by this technique. Both of these have extended $N = 4$ supersymmetry because of usual ($N = 1$) supersymmetry in 10 dimensions. The first one is the Yang–Mills supersymmetry theory for any gauge group G, it has been shown [5] to have vanishing β-function up to 2-loop order, this suggests extra-symmetries. The second model is the SU(4) supergravity theory [6], its equations of motion are the same as those of SO(4) theory, the SO(6) ~ SU(4) algebra of symmetries comes naturally from the 6 dimensional internal structure and is realized in the lagrangian. There is a further invariance of the equations of motion, it is a SU(1, 1) group of transformations realized non-linearly on the scalar fields [7].

Now the theory of ref. [3] has analogous promising features, it is polynomial (up to metric coefficients required for covariance), it admits a large group of invariances: GL(11, **R**) global, included in the local coordinate reparametrization group, SO(1, 10) local thanks to auxiliary non propagating fields $\omega_M{}^{AB}$, 55 local commuting one-parameter groups corresponding to the gauge fields A_{MNP}, and local supersymmetry invariance $N=1$ in 11 dimensions. We are interested in the four-dimensional reduction for phenomenological purposes and because one expects better ultra-violet behaviour there. It was shown in [3] that the reduced theory has the same field content as both the $N = 7$ and $N = 8$ extended supergravity theories. It must have $N = 8$ local supersymmetry invariance but has only obvious O(7) local and GL(7, **R**) global invariances, hence it is clear that one must restore the O(8) symmetry of the lagrangian after reduction, keeping in mind that SU(N) or U(N) symmetries usually hold for massless theories [8]. The larger the symmetry group, the easier it will be to make a careful phenomenological analysis, for some attractive properties of the physics see [9]. Invariances will also play a crucial role in the quantum theory, and for its eventual renormalizability.

Let us first sketch the reduction procedure. The local O(1, 10) invariance of 11 dimensional supergravity can be used to put the moving frame $e_M{}^A$ in block triangular form; $e_m{}^\alpha = 0$; α, a, A (μ, m, M) denote flat (curved) indices between 0 and 3, 4 and 10, i.e. "internal", or general between 0 and 10. This operation preserves local O(7) invariance. GL(7, **R**) invariance will also hold in the sector with no internal coordinate dependence. As shown in [3] the spectrum coincides with the one of ref. [1], we obtain 35 pseudo scalars A_{mnp} but 28 + 7 scalars coming in 2 disjoint representations of O(7) [10], this means that the SO(7) of $N = 7$ supergravity is distinct from our obvious O(7). After a Weyl transformation which eliminates the factor det $[e_m{}^a]$, duality transformations [2] can be done in four dimensions: the 7 scalars φ^i are classically equivalent to the tensor fields $F_{\mu\nu\rho i}$, the 21 abelian vectors are replaced by their "magnetic" duals $B_\mu{}^{ij}$, and together with $B_\mu{}^i = -\frac{1}{2} e_\mu{}^a e_a{}^i$ they will form the representation 28 of SO(8) and gauge 28 U(1) groups. At this stage the lagrangian has become a monster: by duality non polynomial terms in the pseudoscalar fields have appeared.

Let us recall that Majorana spinors in eleven dimensions have 32 components and split, after reduction, into 8 four-component Majorana spinors. These internal indices between 1 and 8 will be written in capital letters A, B ... also (but they appear only in 4 dimensions), they are SU(8) indices [8]. The eleven Γ matrices can be chosen in a factorized form: $\gamma^\alpha \otimes \mathbf{1}_A{}^B$ and $\gamma^5 \otimes \Gamma_A{}^a{}^B = \Gamma^a{}_{(11\,\mathrm{dim.})}$, and the 28 (8 by 8 matrices) $\Gamma^a, \frac{1}{2}[\Gamma^a, \Gamma^b]$ form the chiral spinor representation of the algebra SO(8). Appropriate linear combinations of ψ_μ and ψ_m have the canonical kinetic terms and form an octet $\psi_{\mu A}$ and the 56-plet of spin $\frac{1}{2}$ χ_{ABC} (cf. [1]).

The form of the SO(8) algebra suggests a unification of the 7-bein $e_a{}^i$ with the $7\varphi^i$'s. It can be done by enlarging GL(7, **R**) to SL(8, **R**) and local SO(8). This operation will be used once more, so let us formulate it carefully. ($e_a{}^i$, φ^i) parametrize a matrix of SL(8, **R**): $v_a{}^i$ (indices now run between 4 and 11), but the physical degrees of freedom are described by the "metric" $g^{ij} = v^t \eta v$, $\eta = \mathrm{diag}(-1, 8 \text{ times})$; superfluous components are eliminated by a *local* invariance SO(8) acting on the left of v. Using 7 of these invariances (preserving SO(7)) one can set 7 components of $v_a{}^i$ to zero and arrive at $e_a{}^i$ and φ^i. Mathematically, the scalars live on the homogeneous space SL(8**R**)/SO(8), the physicists are used to coset spaces in the context of gravitation and non linear chiral lagrangians. It is tantalizing to try a generalization including the pseudoscalars. The simplest way to find it is to guess an invariance SU(8)[‡1] extending SO(8) for the full theory and to look for a group containing SL(8, **R**) and SU(8), such that its coset space $S = G/\mathrm{SU}(8)$ be 70-dimensional; 63 + 70 = 133: here is the exceptional group E_7 of Killing–Cartan [11], coming in the theory quite naturally. Like the φ^i's the A_{ijk} appear only polynomially, the reason being also that they parametrize S in a triangular form. To be pedantic this is an Iwasawa decomposition of a non compact real form of E_7 with respect to its maximal compact subgroup SU(8). Concretely, the 11-dimensional theory led to a simple parametrization of the real representation 56 of E_7:

$$\mathcal{V} = \exp P \begin{bmatrix} v^{[c}{}_i v^{d]}{}_j & 0 \\ 0 & v_{[c}{}^i v_{d]}{}^j \end{bmatrix}, \quad P = \begin{bmatrix} 0 & {}^*A^{abcd} \\ A_{abcd} & 0 \end{bmatrix}, \tag{1}$$

with

$${}^*A^{abcd} = \tfrac{1}{24}\,\epsilon^{abcdefgh} A_{efgh} = \tfrac{1}{6}\,\epsilon^{abcdefg}_{(7\,\mathrm{dim.})} e_e{}^m e_f^n e_g^p A^{(7\mathrm{dim.})}_{mnp},$$

$$v_c{}^i = e^{1/4}\begin{bmatrix} e_c{}^i & 0 \\ -\varphi^i e^{-1}, & e^{-1} \end{bmatrix} \text{ the achtbein} \quad \text{and } e = \det[e_i^c]$$

The positive definite scalar lagrangian boils down to:

$$L^0 = -\frac{\sqrt{g_{\mu\nu}}}{192K^2}\,\mathrm{Tr}\,(\partial_\mu \mathcal{G}\cdot\partial^\mu \mathcal{G}^{-1}) \text{ with } \mathcal{G} = \mathcal{V}^{\mathrm{t}}\cdot\mathcal{V}. \tag{2}$$

The group E_7 acts on the right of $\mathcal{V}$ and the "local" SU(8) eliminating the superfluous degrees of freedom, on the left (like GL(4, **R**) and SO(1, 3) for the vierbein $e_\alpha{}^\mu$). E_7 also acts on the fundamental representation 56 $\mathcal{F}_{\mu\nu}$ formed with the field strengths: $G_{\mu\nu}{}^{ij} = \partial_\mu B_\nu{}^{ij} - \partial_\nu B_\mu{}^{ij}$ and the 28, $H_{\mu\nu ij} = \tilde{G}_{\mu\nu}{}^{ij}$ + non-linear terms, appearing in the equations of motion, in such a way that the latter are unified with the Bianchi identities in the single equation $\partial_\mu \tilde{\mathcal{F}}^{\mu\nu} = 0$, where $\tilde{\mathcal{F}}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}\mathcal{F}_{\rho\sigma}$. Non trivial dynamics is generated by the covariant constraint:

$$\Omega\cdot\mathcal{V}\cdot\tilde{\mathcal{F}} = \mathcal{V}\cdot\mathcal{F}, \tag{3}$$

where Ω is the symplectic form invariant under E_7 [11]. It is interesting to note that all E_7 symmetries (including electric-magnetic dualities) can be realized at the lagrangian level by adding 56 abelian potentials $\mathcal{A}'_\mu$ in $L^1 = \frac{1}{4}\mathcal{A}'_\nu\cdot\Omega\cdot\partial_\mu\tilde{\mathcal{F}}^{\mu\nu}$; elimination of 28 of them leads to the usual dissymmetric lagrangians. But quantization is often simpler in "first-order formalism" hence this new result might be useful beyond the classical approximation in any such "dual" theory.

The analogy with general relativity extends to the fermionic Lagrangian: the fermionic fields $\psi_{\mu A}$ and χ_{ABC} transform linearly under the local SU(8), all their couplings to the bosonic fields involve the moving frame $(e_\alpha{}^\mu, \mathcal{V})$ and constant Clebsch–Gordan coefficients: the γ^α and Γ^a matrices. Let us first consider the couplings to the scalar fields alone. In analogy with the connection fields $\omega_{\mu\alpha}{}^\beta$ we can introduce auxiliary (non-propagating) fields $\Omega_{\mu A}{}^B$, they gauge SU(8) but have no kinetic terms. In the following D_μ will denote the covariant derivative in the appropriate representation of SU(8) (28_{C} for $\mathcal{V}$). In first order formalism L^0 becomes:

$$L'^0 = \frac{\sqrt{g_{\mu\nu}}}{48K^2}\,\mathrm{Tr}\,[(D_\mu\mathcal{V}\cdot\mathcal{V}^{-1})\cdot(D^\mu\mathcal{V}\cdot\mathcal{V}^{-1})], \tag{4}$$

and these fermionic terms L_{F} can be rewritten as:

$$L_{\mathrm{F}} = \frac{\epsilon^{\mu\nu\rho\sigma}}{2}\,\bar{\psi}_{\mu A}\gamma_\sigma\gamma_5 D_\nu\psi_{\rho A} + \frac{\mathrm{i}\sqrt{g_{\mu\nu}}}{24}\,\bar{\chi}_{ABC}\bar{\gamma}^\nu D_\nu\chi_{ABC}$$

$$+ \frac{\sqrt{g_{\mu\nu}}}{192}\,\bar{\psi}_{\mu D}\gamma^\nu\gamma^\mu P_\nu^{ABCD}\chi_{ABC}\,. \tag{5}$$

Here D_μ includes the Lorentz connection $\omega_{\mu\alpha}{}^\beta$ in the spin 1/2 representation, and $P_\nu{}^{ABCD}$ is the component of $\partial_\nu\mathcal{V}\cdot\mathcal{V}^{-1}$ perpendicular to SU(8), i.e. the propagating part of the scalar fields. The other component drops out of L'_0 after elimination of $\Omega_{\mu A}{}^B$, thanks to SU(8) invariance.

The fermionic couplings $K\Theta(G)$ of the fields $G_{\mu\nu}{}^{ij}$ are found by reduction in L_{R}, the second order lagrangian; they respect the E_7 and SU(8) invariances if one uses the first order formalism. Indeed the field strengths and $H_{\mu\nu ij} = -4(\delta\tilde{L}_R/\delta G^{ij}_{\mu\nu})$, i.e. $\mathcal{F}_{\mu\nu}$ still belong to a 56-plet of E_7, but (3) no longer holds. However, one can add uniquely to this multiplet another one, quadratic in the fermion fields, such that the new fields $(\hat{G}_{\mu\nu}{}^{ij}, \hat{H}_{\mu\nu ij}) = \hat{\mathcal{F}}$ satisfy the duality constraint (3). L_1 acquires an additional E_7 invariant term:

[‡1] U(8) is suggested by ref. [8], however the self-duality condition of ref. [1]:
$A_{ijkl} + \mathrm{i}B_{ijkl} = z_{ijkl} = (1/4!)\,\epsilon_{ijklmnop}\bar{z}^{mnop}$
reduces it to SU(8).

$$L'^{1} = \tfrac{1}{4}\mathcal{A}'_{\nu}\cdot\Omega\cdot\partial_{\mu}\tilde{\mathcal{F}}^{\mu\nu} + \tfrac{1}{2}K\Theta(\hat{\mathcal{F}})\,, \tag{6}$$

where

$$32\,\Theta/\sqrt{g_{\mu\nu}} = \bar{\psi}_{\mu A}\gamma^{[\nu}\hat{\mathcal{F}}_{AB}\gamma^{\mu]}\psi_{\nu B} - \tfrac{1}{2}\,\mathrm{i}\bar{\psi}_{\mu C}\hat{\mathcal{F}}_{AB}\gamma^{\mu}\chi_{ABC}$$
$$-(\eta/144)\epsilon^{ABCDEFGH}\bar{\chi}_{ABC}\hat{\mathcal{F}}_{DE}\chi_{FGH}\,, \tag{7}$$

where $\eta = \pm 1$ [12] depends on the explicit representation of the real Clifford algebra Γ^{a} ($a = 1, \dots 7$). We have used the SU(8) covariant tensor $\hat{\mathcal{F}}_{AB}$ constructed from $\hat{\mathcal{F}}_{\mu\nu}$ by:

$$\mathcal{F}_{AB} = ((\Gamma_{rs})_{AB}\,;\;\; \mathrm{i}\gamma_5(\Gamma^{rs})_{AB})\mathcal{V}\cdot\mathcal{F}_{\mu\nu}\gamma^{\mu\nu}\,. \tag{8}$$

To summarize, the lagrangian of SO(8) supergravity admits global E_7 and local SU(8) invariances and reads:

$$L = -\frac{\sqrt{g_{\mu\nu}}}{4K^2}R(e_{\mu}{}^{\alpha}, \omega_{\mu}{}^{\alpha}{}_{\beta}) + L'_1 + L'_0 + L_{\mathrm{F}} +$$
$$+ \text{eventual quartic terms}\,. \tag{9}$$

We have two constraints: $\mathcal{V}$ is a matrix representing E_7 in the fundamental representation and (3) holds for $\hat{\mathcal{F}}$; both are compatible with the symmetries. We believe that as in $N = 4$ supergravity all quartic terms can be absorbed in supercovariant expressions (for example $\hat{\mathcal{F}}$ appearing in (7)), or generated by the elimination of auxiliary fields like $\Omega_{\mu A}{}^{B}$. It is pointless to give them without describing the supersymmetry algebra [13].

In order to reproduce the results of [1], one must choose a specific parametrization of the coset space $E_7/SU(8)$ (i.e. a choice of SU(8) gauge). It coincides only to first order in the scalar fields with the symmetric gauge $\mathcal{V} = \mathcal{V}^{t}$ where $\mathcal{V}$ is generated by the Lie generators perpendicular to SU(8), we have called it the "analytic gauge". Details of the reduction procedure and important, if sophisticated, algebraic identities will be the object of a longer article [12].

In this letter we have worked our way from O(7) to E_7, without paying much attention to supersymmetry, it is conjectured that the shrinking of the lagrangian accompanying the growth of the group will continue when one uses supercovariance. Several generalizations must be investigated further, first of all the use of other non-compact symmetry groups without coupled ghosts, then the introduction of a non-vanishing gauge coupling and the associated potentials, and also the quantized version of this theory. Indeed the non-trivial topology of our phase space suggests a large number of non perturbative effects, which would be used to cure some of the phenomenological difficulties of the classical model [9]. One recalls that in 2 dimensions the vector auxiliary field of the non-linear σ-model may become dynamical [13].

We are grateful to Joel Scherk for his collaboration to the first computations of this work.

References

[0] An extensive list of references can be found in: S. Ferrara, Erice Lectures (1978), CERN TH-2514.

[1] B. De Wit and D.Z. Freedman, Nucl. Phys. B130 (1977) 105.

[2] F. Gliozzi, J. Scherk and D. Olive, Nucl. Phys. B122 (1977) 253;
L. Brink, J.H. Schwarz and J. Scherk, Nucl. Phys. B121 (1977) 77.

[3] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.

[4] W. Nahm, Nucl. Phys. B135 (1978) 149.

[5] D.R.T. Jones, Phys. Lett. 72B (1977) 199;
E.C. Poggio and H.N. Pendleton, Phys. Lett. 72B (1977) 200.

[6] E. Cremmer, J. Scherk and S. Ferrara, Phys. Lett. 74B (1978) 61.

[7] E. Cremmer, J. Scherk, Phys. Lett. 74B (1978) 341.

[8] R. Haag, J.T. Lopuszanski, M. Sohnius, Nucl. Phys. B88 (1975) 257.

[9] M. Gell-Mann, Lecture at the Washington meeting of the APS (April 1977).

[10] We thank D. Freedman for bringing this fact to our attention.

[11] E. Cartan, Oeuvres Complètes T1, Part I (Thèse) (Gauthier–Villars, Paris, 1952).

[12] E. Cremmer and B. Julia, to be published.

[13] A. D'Adda, M. Lüscher and P. Di Vecchia, Nordita preprint;
E. Witten, Harvard preprint.

DYNAMICS OF DIMENSIONAL REDUCTION ☆

Peter G.O. FREUND and Mark A. RUBIN
Enrico Fermi Institute and Department of Physics, The University of Chicago,
Chicago, IL 60637, USA

Received 4 August 1980
Revised manuscript received 19 September 1980

In d-dimensional unified theories that, along with gravity, contain an antisymmetric tensor field of rank $s-1$, preferential compactification of $d-s$ or of s space-like dimensions is found to occur. This is the case in 11-dimensional supergravity where $s = 4$.

Kaluza's old idea [1] of dimensional reduction has been revived in the context of nonabelian gauge theories [2] and of extended supergravity [3,4]. One essentially accounts for a seemingly complicated theory with many force and matter fields in a low-dimensional space–time, in terms of a simple geometrical theory in a space–time of higher dimensionality. The extra dimensions are assumed to compactify with a very small characteristic size. Attempts to justify this compactification have included further Bose fields in the original (higher-dimensional) theory and observed the appearance of classical solutions in which some of the space dimensions "spontaneously" compactified [5]. Such complication of the original theory is contrary to the spirit of unification. Here we remark that bosonic fields other than gravity, automatically appear in higher-dimensional supergravity theories [3], and that they can induce spontaneous compactification. The number of compactified dimensions can then take only two preferential values. For 11-dimensional supergravity we find that either 7 or 4 space-like dimensions compactify. In the first case ordinary "large" space–time has therefore 1 time and 3 space dimensions, a pleasing result.

To start, consider first Einstein–Maxwell theory in d dimensions. The field equations are (throughout this paper we adopt the notations and conventions of ref. [6], except that Greek indices run from 1 to d)

$$R^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu} R = -8\pi G \theta^{\mu\nu} , \tag{1a}$$

$$(1/\sqrt{|g|})\partial_\mu(\sqrt{|g|} F^{\mu\nu}) = 0 , \tag{1b}$$

$$\theta^{\mu\nu} = F_\rho{}^\mu F^{\rho\nu} - \tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} g^{\mu\nu} . \tag{1c}$$

$F^{\mu\nu}$ being a totally antisymmetric tensor of rank two, we look for solutions of these equations such that (a) d-dimensional space–time becomes the product of a two-dimensional and of a $(d-2)$-dimensional Riemann manifold: $M = M_2 \times M_{d-2}$, and (b) $F^{\mu\nu}$ is proportional to the totally antisymmetric Levi-Cività tensor on M_2. In other words, the metric is then of the form

$$g_{\mu\nu} = \begin{pmatrix} g_{mn}(x^p) & 0 \\ 0 & g_{\bar{m}\bar{n}}(x^{\bar{p}}) \end{pmatrix} , \tag{2}$$

with m, n, p taking two values and $\bar{m}, \bar{n}, \bar{p}$ the remaining $d-2$ values. The Maxwell equations admit the solution

$$F^{\mu\nu} = (\epsilon^{\mu\nu}/\sqrt{|g_2|}) f , \tag{3a}$$

where

$$\begin{aligned} \epsilon^{\mu\nu} &= \epsilon^{mn} , \quad && \text{for } \mu = m, \nu = n , \\ &= 0 , && \text{otherwise} . \end{aligned} \tag{3b}$$

Here $\epsilon^{mn} = -\epsilon^{nm}$ and say $\epsilon^{mn} = +1$ for $m < n$, g_2 is the determinant of the 2×2 matrix g_{mn}, and f is a constant with dimension mass squared. For the solution (2), (3), the second term in $\theta^{\mu\nu}$ becomes an overall cos-

---

☆ Work supported in part by the NSF, Grant No. PHY79-23669.

mological term, while the first term vanishes except for $\mu = m$ and $\nu = n$ when it is also proportional to g^{mn}. The source term in the Einstein equations thus reduces to cosmological terms with different cosmological constants on M_2 and M_{d-2}.

Defining the scalar curvatures $R_2 = g^{mn}R_{mn}$ and $R_{d-2} = g^{\bar{m}\bar{n}}R_{\bar{m}\bar{n}}$ of M_2 and M_{d-2}, we have $R = R_2 + R_{d-2}$. The Einstein equations yield

$$R_{d-2} = \lambda\,, \quad R_2 = -2\,\frac{d-3}{d-2}\,\lambda\,, \tag{4a}$$

with

$$\lambda = 8\pi G f^2 \,\mathrm{sgn}(g_2) \tag{4b}$$

[$\mathrm{sgn}(g_2) = g_2/|g_2|$ appears in the expression of $\theta^{\mu\nu}$ in terms of the $F^{\mu\nu}$ of eqs. (3)].

In our notation a compact Riemann manifold has negative scalar curvature. For $d \geqslant 4$, it follows from $Gf^2 > 0$ that R_2 and R_{d-2} have opposite signs while R_{d-2} and g_2 have the same sign. Thus when the time dimension is in M_2 then M_{d-2} is compact and vice versa.

We can now replace the Maxwell vector field by an antisymmetric tensor field of rank s–1. The corresponding field strengths will then be the components of a completely antisymmetric tensor $F^{\mu_1 \cdots \mu_s}$ of rank s (we refrain from the use of differential forms). The energy–momentum tensor, and the field equations for F become then:

$$\theta^{\mu\nu} = F_{\alpha_1 \ldots \alpha_{s-1}}{}^{\mu} F^{\alpha_1 \ldots \alpha_{s-1}\,\nu} - \frac{1}{2s} F_{\alpha_1 \ldots \alpha_s} F^{\alpha_1 \ldots \alpha_s} g^{\mu\nu}\,, \tag{5a}$$

and

$$(1/\sqrt{|g|})\,\partial_\mu (\sqrt{|g|}\, F^{\mu\alpha_2 \cdots \alpha_s}) = 0\,. \tag{5b}$$

The indices m, n, p $(\bar{m}, \bar{n}, \bar{p})$ in eq. (2a) now run over s $(d-s)$ values and the analogue of the solution (3) is

$$F^{\mu_1 \cdots \mu_s} = (\epsilon^{\mu_1 \cdots \mu_s}/\sqrt{|g_s|})\, f\,, \tag{6a}$$

$$\begin{aligned} \epsilon^{\mu_1 \cdots \mu_s} &= \epsilon^{m_1 \ldots m_s}\,, && \text{for } \mu_1 = m_1, \ldots, \mu_s = m_s\,, \\ &= 0\,, && \text{otherwise}\,. \end{aligned} \tag{6b}$$

Defining R_s and R_{d-s} by analogy with R_2 and R_{d-2} we then find

$$R_{d-s} = \frac{(s-1)(d-s)}{d-2}\,\lambda\,, \quad R_s = -\frac{s(d-s-1)}{d-2}\,\lambda\,, \tag{7a}$$

$$\lambda = 8\pi G f^2 \,\mathrm{sgn}(g_s)\,. \tag{7b}$$

For $d > s + 1$ again either s or d–s space-like dimensions compactify. The solutions (6) exist *only* for such $M_s \times M_{d-s}$ splittings of the original d-dimensional manifold (while completing this work we received a paper by Aurilia et al. [7], in which the special case $d = 4, d-s = 0, M \equiv M_s$ of the solution (6) is considered in a different context).

Without supersymmetry, antisymmetric tensor fields of arbitrary rank ($1 \leqslant s \leqslant d$) can be introduced, and nothing can be said about the number of compactified dimensions. In $d = 11$ supergravity, supersymmetry forces the introduction of one such field with $s = 4$. Solutions of the form (6) exist in this case. The corresponding fermion fields vanish as does the self-coupling of the antisymmetric tensor field of form (6). Since $d = 11 > s + 1 = 5$, this leads to the preferential compactification of 7 or 4 space-like dimensions, so that physical (i.e., "large") space–time must be either 4 or 7 dimensional.

The question becomes now: why to start with one time and ten space dimensions in the first place? Here are a few remarks about this. The number of time-like dimensions cannot exceed one, or space–time would contain closed time-like curves. Why then the large number of space dimensions? The divergences of supersymmetric theories are milder than those of ordinary non-supersymmetric ones [8]. In 11 space–time dimensions $N = 1$ supergravity may lead to a convergent quantum theory (possibly on account of the large number of Fermi dimensions involved). Beyond eleven dimensions higher "spin" fields must enter [4], and to the extent that such fields cannot be consistently brought to interact, 11-dimensional supergravity is also a "maximal" theory. The requirements of causality and quantum consistency may then help pin down the dimensionality of space–time and the dynamics of the theory. Finally, 11 is the only space–time dimension such that no matter (Yang–Mills) supermultiplets exist, so that with one graviton the basic fields are unambiguously determined.

Concerning the size of the compactified dimensions, on dimensional grounds we expect $f \sim \mathrm{O}(G^{-1})$, so that the compactification size is of order of the Planck length.

But then ordinary space–time gets its own cosmological constant of opposite sign but again of order G^{-1} [eqs. (7)]. This is immense, so how does it get removed? The matter fields produced by dimensional reduction, and quantum effects could all contribute to the cosmological term, but how such a drastic cancellation occurs without affecting our classical arguments above is unclear. Moreover, the dimensionally reduced theory in turn contains many gauge and scalar Bose fields. Do these lead to further compactification of dimensions, and if yes, are the characteristic sizes sufficiently large to agree with cosmological evidence?

We wish to thank Dr. Thomas L. Curtright for a valuable discussion.

References

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. 966 (1921);
O. Klein, Z. Phys. 37 (1926) 895.

[2] B. de Witt, Dynamical theories of groups and fields (Gordon and Breach, New York, 1965) p. 139;
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Rep. Math. Phys. 1 (1970) 29;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711.

[3] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.

[4] W. Nahm, Nucl. Phys. B135 (1978) 149;
F. Gliozzi, J. Scherk and D. Olive, Nucl. Phys. B122 (1977) 253.

[5] E. Cremmer and J. Scherk, Nucl. Phys. B108 (1976) 409; B118 (1977) 61;
J.F. Luciani, Nucl. Phys. B135 (1978) 111.

[6] S. Weinberg, Gravitation and cosmology (Wiley, New York, 1972).

[7] A. Aurilia, M. Nicolai and P.K. Townsend, preprint Ref. TH 2875-CERN.

[8] See P. Fayet and S. Ferrara, Phys. Rep. 32C (1977) 249.

SPONTANEOUS COMPACTIFICATION OF ELEVEN-DIMENSIONAL SUPERGRAVITY *

F. ENGLERT [1]
CERN, Geneva, Switzerland
and Faculté des Sciences, Université Libre de Bruxelles, Belgium

Received 31 August 1982

Three distinct geometries on S_7 arise as solutions of the classical equations of motion in eleven dimensions. In addition to the conventional riemannian geometry, one can also obtain the two exceptional Cartan–Schouten compact flat geometries with torsion. Possible implications of these results are mentioned.

Simple supergravity in eleven dimensions [1] may turn out to be of more fundamental significance than the ungauged version of the maximal ($N = 8$) extension of four-dimensional supergravity [2]; in the trivial dimensional reduction process leading from the former theory to the latter, relevant information may well have been lost. Following this line of thought, it is of interest to look for possible non-trivial classical solutions in eleven dimensions which would describe a compactification of seven dimensions into a sphere S_7 and which would keep four space-time dimensions non-compact. Indeed a Kaluza–Klein dimensional reduction [3] around $S_7 \equiv SO(8)/SO(7)$ would lead to an SO(8) gauge theory which could be compared and perhaps identified with the gauged version of the $N = 8$ theory [4]: in this way one would learn whether or not this gauged version is also contained in the eleven-dimensional theory ‡1.

In this paper we shall derive the surprising result that there exists not only one, but three geometries on S_7 which emerge as solutions of the equations of motion in eleven dimensions. In addition to the "trivial" riemannian geometry on S_7, one also obtains the two "non-trivial" geometries for which the linear connection in S_7 contains just the torsion terms required to define an absolute parallelism on the seven sphere. This property of the eleven-dimensional action is particularly remarkable because, apart from group spaces, S_7 is the only parallelizable compact manifold, a feature linked to the fact that it is the space of unit octonions [7].

We now present and briefly discuss these solutions. We defer to a separate publication the detailed analysis of the supersymmetry and gauge symmetries they would induce in the four-dimensional space–time background.

The bosonic action of the theory is

$$S = \int \sqrt{g^{(11)}}\, \mathrm{d}^{(11)}x \,\{-\tfrac{1}{2}R - \tfrac{1}{48}F_{MNPQ}F^{MNPQ}$$

$$+ [\sqrt{2}/6\cdot(4!)^2]\,(1/\sqrt{g^{(11)}})\,\epsilon^{M_1M_2\ldots M_{11}}$$

$$\times F_{M_1M_2M_3M_4}F_{M_5M_6M_7M_8}A_{M_9M_{10}M_{11}}\}. \qquad (1)$$

We use the metric (+, — — ..), $\epsilon^{0123\ldots} = +1$ and $F_{MNPQ} = 4!\,\partial_{[M}A_{NPQ]}$. The field equations are

$$R_{MN} - \tfrac{1}{2}g_{MN}R =$$

$$= -\tfrac{1}{48}(8F_{MPQR}F_N{}^{PQR} - g_{MN}F_{SPQR}F^{SPQR}), \qquad (2)$$

$$F^{MNPQ}{}_{;M} = -[\sqrt{2}/2\cdot(4!)^2]\,(1/\sqrt{g^{(11)}})\,\epsilon^{M_1\ldots M_8NPQ}$$

$$\times F_{M_1M_2M_3M_4}F_{M_5M_6M_7M_8}. \qquad (3)$$

---

* Supported in part by the Belgian State under the contract ARC 79/83-12.

[1] Permanent address: Pool de Physique, Université Libre de Bruxelles, Campus Plaine, CP 225 Bd. du Triomphe, 1050 Bruxelles, Belgium

‡1 This possibility was already suggested by Duff and Pope [5] (see Duff and Toms [6]).

We look for solutions of (2) and (3) which describe a direct product of S_7 with a four-dimensional Einstein space to cope with an eventual cosmological constant. Hence we must have

$$R_{mn} = \gamma g_{mn} \quad (m, n = 4, 5, 6, 7, 8, 9, 10), \tag{4}$$

$$R_{\mu\nu} = \gamma' g_{\mu\nu} \quad (\mu, \nu = 0, 1, 2, 3), \tag{5}$$

$$R_{m\mu} = 0, \tag{6}$$

where γ and γ' are constants and $\gamma < 0$.

Eq. (3) is consistent with the existence of a non-vanishing expectation value for the auxiliary field $F^{\mu\nu\sigma\lambda}$, namely [8]

$$F^{\mu\nu\sigma\lambda} = (1/\sqrt{-g^{(4)}})\,\epsilon^{\mu\nu\sigma\lambda}\, f/(4!)^{1/2},$$

$$f = \text{constant}. \tag{7}$$

If we assume that no other components of F^{MNPQ} acquire a non-vanishing expectation value, we see that (4), (5) and (6) solve eq. (2) provided

$$f^2 = -F_{\mu\nu\sigma\lambda}F^{\mu\nu\sigma\lambda} = -72\gamma, \quad \gamma' = -2\gamma. \tag{8, 9}$$

Thus the "trivial" S_7 is indeed a solution for arbitrary γ, or equivalently arbitrary radius.

Let us now admit non-vanishing expectation values for the four-scalars F^{mnpq}. The solution (7) remains valid, but (3) now yields an additional equation:

$$F^{mnpq}{}_{;m} = [\sqrt{2}/(4!)^{3/2}]\, f(1/\sqrt{-g^{(7)}})\,\epsilon^{npqrstu}F_{rstu}. \tag{10}$$

It will turn out that the solutions of eq. (10) are related to the torsion tensors required to parallelize S_7. We shall therefore first rederive the Cartan–Schouten equations which these tension tensors must satisfy.

A Riemann space is parallelizable if, by adding to the riemannian symmetric connection $\Gamma^m{}_{np}$ a suitable totally antisymmetric torsion tensor $S^m{}_{np}$ ($S_{mnp} = S_{[mnp]}$), one can "flatten" the space in the sense that

$$R^m{}_{npq}(\{\Gamma^s{}_{tu} + S^s{}_{tu}\}) = 0. \tag{11}$$

Using the cyclic identities for $R^m{}_{npq}(\{\Gamma^s{}_{tu}\})$ one easily obtains the necessary and sufficient conditions for such a torsion, and hence for a parallelism, to exist:

$$\overset{0}{R}_{mnpq} = S_{tmn}S^t{}_{pq} - S_{t[mn}S^t{}_{p]q}, \tag{12}$$

$$S_{mnp;q} = S_{t[mn}S^t{}_{p]q} = S_{[mnp,q]}. \tag{13}$$

Here and in what follows $\overset{0}{R}_{mnpq} \equiv R_{mnpq}(\{\Gamma^s{}_{tu}\})$ and the covariant derivatives are taken with respect to the Riemann connection $\Gamma^s{}_{tu}$; the last equality in (13) follows from the antisymmetry of $S_{t[mn}S^t{}_{p]q}$ in the *four* indices m, n, p and q. Contracting any two indices in (12) with g^{rs} and $S^{rs}{}_t$ and using the explicit form of $\overset{0}{R}_{mnpq}$ for S_7, namely

$$\overset{0}{R}_{mnpq} = \tfrac{1}{6}\gamma(g_{mp}g_{nq} - g_{mq}g_{np}), \tag{14}$$

one gets

$$S^{tr}{}_m S_{trn} = \gamma g_{mn}, \tag{15}$$

$$S_{tm}{}^r S_{rn}{}^s S_{sp}{}^t = \tfrac{1}{2}\gamma\, S_{mnp}. \tag{16}$$

Eqs. (15) and (16) were first obtained by Cartan and Schouten [7] who showed that their solutions define two types of parallelisms on S_7.

We shall show that each parallelism is described by a torsion $\underset{+}{S}_{mnp}$ or $\underset{-}{S}_{mnp}$ which satisfy different linear partial differential equations. To this effect, consider the algebraic equations

$$X^{mnp} \equiv X^{[mnp]} = \pm(6/-\gamma)^{1/2}(1/4!)(1/\sqrt{-g^{(7)}})$$

$$\times\, \epsilon^{mnpqrst}X_{u[rs}X^u{}_{t]q}. \tag{17}$$

Performing the ϵ algebra, one may verify that both eqs. (17) admit solutions satisfying (12) and which of course also satisfy (15) and (16). These solutions determine the torsions $\underset{+}{S}$ and $\underset{-}{S}$ up to a sign which can be fixed by (13). Using the latter equation in the right-hand side of (17) one gets

$$\underset{+}{S}^{mnp} = +(6/-\gamma)^{1/2}(1/4!)(1/\sqrt{-g^{(7)}})$$

$$\times\, \epsilon^{mnpqrst}\underset{+}{S}_{[rst,q]}, \tag{18}$$

$$\underset{-}{S}^{mnp} = -(6/-\gamma)^{1/2}(1/4!)1/(\sqrt{-g^{(7)}})$$

$$\times\, \epsilon^{mnpqrst}\underset{-}{S}_{[rst,q]}. \tag{19}$$

These "fields" satisfy $\underset{\pm}{S}^{mnp}{}_{;m} = 0$; they may be interpreted as "potentials" which, in this "Landau gauge", are dual (or anti-dual) to their field strength

$\underset{+}{S}_{[mnp,q]}$ ($\underset{-}{S}_{[mnp,q]}$). These potentials are eigenfunctions of the Laplace operator on S_7:

$$\underset{\pm}{S}^{mnp;q}{}_{;q} = -\tfrac{2}{3}\gamma \underset{\pm}{S}^{mnp}. \tag{20}$$

Eq. (20) follows from (13) using (15) and (16); it may also be obtained directly from (18) or (19) using $\underset{\pm}{S}_{[rst,q]} = \underset{\pm}{S}_{rst;q}$.

We now solve eq. (10) by postulating

$$F_{mnpq} = \lambda \underset{+}{S}_{[npq,m]}, \tag{21}$$

or

$$F_{mnpq} = \lambda \underset{-}{S}_{[npq,m]}, \tag{22}$$

where λ is a constant. Indeed, using (18), (19) and (20) we can satisfy (10) provided

$$\pm f = \sqrt{-\gamma}\, 2^{5/2}. \tag{23}$$

As the right-hand side of (23) is positive, one must choose (21) or (22) according to the sign of f.

The expressions (7), (21) [or (22)] and (23) define solutions of eq. (3) if g_{mn} is the metric tensor of S_7. One must still verify (4), (5) and (6). From (2) we see that these equations can be satisfied only if $F_{mpqr}F_n{}^{pqr}$ is proportional to g_{mn}. This is indeed the case as follows from (21) [or (22)] and (18) [or (19)]: inverting (18) [or (19)] and performing the ϵ algebra, we get, using (15)

$$F_{mpqr}F_n{}^{pqr} = \tfrac{2}{3}\lambda^2\gamma^2 g_{mn}. \tag{24}$$

Eq. (2) now yields (4), (5) and (6) and determines λ^2:

$$\lambda^2 = -12/\gamma. \tag{25}$$

Thus we have obtained, in addition to the "trivial" S_7 solution characterized by $F_{mnpq} = 0$, new solutions where the four-scalars take the values (21) or (22) with λ^2 determined by (25). The results corresponding to (8) and (9) are, for these new solutions,

$$f^2 = -F_{\mu\nu\sigma\lambda}F^{\mu\nu\sigma\lambda} = -32\gamma, \tag{26}$$

$$F^2 = F_{mpqr}F^{mpqr} = -56\gamma, \tag{27}$$

$$\gamma' = -\tfrac{5}{3}\gamma, \tag{28}$$

and γ is, of course, still arbitrary. The geometrical significance of (21) and (22) is clear from eqs. (18) and (19): the admissible expectation values of the four-scalars F_{mnpq} are just those for which the potentials A_{npq} in the "Landau gauge" are, up to a scale factor whose absolute value is determined by (25), equal to one of the torsion tensors $\underset{\pm}{S}^{npq}$ which "flattens" the seven-sphere.

A few concluding remarks are in order:

(a) the non-trivial solutions [(26), (27), (28)], as well as the trivial one [(8), (9)], will necessarily lead, after a Kaluza–Klein reduction around $S_7 \equiv SO(8)/SO(7)$, to four-dimensional actions containing SO(8) gauge fields with a coupling constant proportional to $|\gamma|$ in Planck units. What about supersymmetry? It can be shown [5] that, provided four-dimensional space-time is taken to be the anti-de Sitter space satisfying eq. (9), the trivial solution on S_7 preserves the full supersymmetry of the original action (1) with fermions included, but that supersymmetry is broken by the non-trivial solution. Hence the four-dimensional action deducible from the reduction around the trivial S_7 geometry will be fully supersymmetric and this suggests of course that it might describe the gauged version of $N = 8$ supergravity [4]. On the other hand, the reduction around the non-trivial S_7 geometries will reveal a spontaneous breakdown of supersymmetry [‡2].

(b) The appealing mathematical properties of the non-trivial S_7 solutions are marred by the reappearance of a cosmological constant [eq. (28)] which was already there in the trivial S_7 solution: thus the "flattening" of S_7 is not accompanied by a vanishing of the cosmological constant. However, the value of the measurable cosmological constant can anyhow not be predicted without understanding the effect of the quantum fluctuations which are not considered here.

(c) Finally, we have not at all discussed the relative stability of the different classical solutions. There may be no sensible way of attacking this problem as long as the physical implications of the "wrong" sign induced by the conformal part of gravity are not mastered. This may require the link between field theory and cosmology provided by the conjecture that the path integral over metrics is stabilized by the entropy stored in a "foam of universes" [9]. In this respect, it is perhaps encouraging to notice that if eleven-dimensional supergravity is physically relevant, and if spontaneous compactification does take place, a multiplicity of "universes" will already be unavoidably generated at the level of "multi-soliton" solutions.

[‡2] I thank M. Duff for pointing out to me that the spontaneous character of this breaking is preserved under the Kaluza–Klein dimensional reduction process.

I thank the members of the Department of Physics and Astronomy of the Tel-Aviv University where part of this work was completed, for their warm hospitality and active interest. In particular I am grateful to T. Banks and A. Casher for illuminating discussions. I am also greatly indebted to R. Brout, M. Duff, B. de Wit, M. Green, H. Nicolai and P. Windey for encouragement and clarifying remarks.

References

[1] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.

[2] E. Cremmer and B. Julia, Phys. Lett. 80B (1978) 48; Nucl. Phys. B159 (1979) 141.

[3] B. de Wit, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965).

[4] B. de Wit and H. Nicolai, Phys. Lett. 108B (1981) 285; CERN preprint TH-3291 (1982).

[5] M.J. Duff and C.N. Pope, to be published; B. de Wit, private commnuication.

[6] M.J. Duff, CERN preprint TH 3232 (1982); M.J. Duff and D.J. Toms, CERN preprint TH 3259 (1982).

[7] E. Cartan and J.A. Schouten, Proc. K. Akad. Wet. Amsterdam 29 (1926) 933.

[8] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233.

[9] A. Casher and F. Englert, Phys. Lett. 104B (1981) 117; F. Englert, Quantum field theory and cosmology, in Fundamental interactions: Cargèse 1981 (Plenum, New York, 1982).

GRAVITY WITH EXTRA GAUGE SYMMETRY ☆

I. BARS and S.W. MacDOWELL
J.W. Gibbs Laboratory, Yale University, New Haven, CT 06511, USA

Received 9 June 1983

In addition to the tangent space Lorentz gauge invariance, we introduce an extra gauge principle in the gauge formulation of gravity, in any dimension. The connection ω^{ij}_μ, which starts out as an elementary field, undergoes a gauge transformation which mixes tangent space indices (i, j) with base indices (μ). The equations of motion determine ω^{ij}_μ as the sum of usual riemannian connection and a completely antisymmetric torsion piece. In 11 dimensions our theory reproduces the bosonic sector of supergravity if we identify torsion with the 3-index gauge potential $A_{\mu\nu\lambda}$. This generalizes Englert's recent observation, since in our formulation torsion = $A_{\mu\nu\lambda}$ for any field configuration, not just a solution.

It is well known that gravity can be described as a gauge theory [1]. To do so, a tangent space is defined at each point of space time by defining a tetrad of vector fields. A vierbein e^i_μ and its inverse e^μ_i are introduced to transform tangent space indices (latin $i, j, \ldots$) to coordinate basis indices (greek $\mu, \nu, \ldots$) and vice versa. In tangent space the metric η_{ij} is flat, while in the coordinate basis the metric is $g_{\mu\nu} = e^i_\mu e_{i\nu}$. The gauge group acts on tangent space indices i, j, etc. and is identified with the Lorentz group in tangent space. The gauge potential ω^{ij}_μ is interpreted as the connection. It is convenient to work with this mixed set of indices for the gauge structure of the theory to become apparent.

Then, the field strength

$$R^{ij}_{\mu\nu} = \partial_\mu \omega^{ij}_\nu - \partial_\nu \omega^{ij}_\mu + [\omega_\mu, \omega_\nu]^{ij} \tag{1}$$

is identified with the curvature, and the antisymmetric covariant derivative of the vierbein

$$T^i_{\mu\nu} = \tfrac{1}{2}\,(\mathrm{D}_\mu e^i_\nu - \mathrm{D}_\nu e^i_\mu)\,,$$

$$\mathrm{D}_\mu e^i_\nu = \partial_\mu e^i_\nu + \omega^{ij}_\mu e_{j\nu}, \tag{2}$$

is identified with the torsion. This identification is independent of any other fields that may interact with gravity. It is the equations of motion with respect to ω^{ij}_μ which eventually determine the connection as a function of the vierbein (riemannian connection) plus any torsion as a function of the matter fields. For example in pure gravity the equation of motion for ω^{ij}_μ reduces to $T^i_{\mu\nu} = 0$, i.e. zero torsion, which, in turn determines $\omega^{ij}_\mu = \overset{\circ}{\omega}{}^{ij}_\mu$ as the usual riemannian connection as a function of the vierbein:

$$\overset{\circ}{\omega}{}^{ij}_\mu = e^{\lambda i} e^{\sigma j}(C_{\mu\lambda\sigma} - C_{\lambda\sigma\mu} - C_{\sigma\mu\lambda}),$$

$$C_{\mu\lambda\sigma} = \tfrac{1}{2}\, e_{k\mu}(\partial_\lambda e^k_\sigma - \partial_\sigma e^k_\lambda). \tag{3}$$

In supergravity or theories with spinning matter the torsion comes out non-zero. For an example of this procedure see ref. [2] and the material that follows.

In this paper we extend this formalism by requiring an extra gauge invariance beyond the local Lorentz invariance of the tangent space. This will be valid in any number of dimensions and has some special features in certain dimensions, such as $d = 7, 11$, etc. Accordingly, both the latin $(i, j, \ldots)$ and greek $(\mu, \nu, \ldots)$ indices will run from 1 to d.

The usual local Lorentz transformations are

$$\delta\omega^{ij}_\mu = \mathrm{D}_\mu \epsilon^{ij}(x) = \partial_\mu \epsilon^{ij} + [\omega_\mu, \epsilon]^{ij},$$

$$\delta e^i_\mu = \epsilon^{ij} e_{j\mu}, \tag{4}$$

under which the curvature and torsion of eqs. (1) and (2) are covariant. We now define the extra gauge transformation

---

☆ Research supported in part by the US Department of Energy under contract No. DE-AC02-76ER03075.

$$\delta\omega_\mu^{ij} = e^{\lambda i}e^{\sigma j}(\partial_\mu\Lambda_{\lambda\sigma} + \partial_\lambda\Lambda_{\sigma\mu} + \partial_\sigma\Lambda_{\mu\lambda}),$$

$$\delta e_\mu^i = 0, \tag{5}$$

where we have used the inverse of the vierbein $e^{\lambda i}$ to transform greek to latin indices, and $\Lambda_{\mu\nu}$ is an antisymmetric tensor.

The curvature and torsion transform in a non-trivial way, but remarkably, an invariant action can be written down:

$$S_0 = -\frac{1}{4}\int \mathrm{d}^d x\,(\det e)(R_{\mu\nu}^{ij}e_i^\mu e_j^\nu + \tfrac{1}{2}A_{\mu\nu\lambda}A^{\mu\nu\lambda}), \tag{6a}$$

where

$$A_{\mu\nu\lambda} = \tfrac{1}{3}(T_{\mu\nu}^i e_{i\lambda} + T_{\nu\lambda}^i e_{i\mu} + T_{\lambda\mu}^i e_{i\nu}) \tag{6b}$$

is the completely antisymmetric projection of torsion. We can easily prove the invariance of this action by using the following transformation laws:

$$\begin{aligned}\delta A_{\mu\nu\lambda} &= \tfrac{1}{3}(\delta\omega_\mu^{ij}e_{i\lambda}e_{j\nu} + \delta\omega_\lambda^{ij}e_{i\nu}e_{j\mu} + \delta\omega_\nu^{ij}e_{i\mu}e_{j\lambda})\\ &= -(\partial_\mu\Lambda_{\nu\lambda} + \partial_\nu\Lambda_{\lambda\mu} + \partial_\lambda\Lambda_{\mu\nu}),\end{aligned} \tag{7a}$$

and

$$\begin{aligned}&\int \mathrm{d}^d x\,(\det e)\,e_i^\mu e_j^\nu\,\delta R_{\mu\nu}^{ij}\\ &= -\int \mathrm{d}^d x\,(\det e)(e_i^\mu e_j^\nu e_k^\lambda + e_i^\nu e_j^\lambda e_k^\mu + e_i^\lambda e_j^\mu e_k^\nu)\\ &\quad\times T_{\mu\lambda}^k\,\delta\omega_\nu^{ij}\\ &= \int \mathrm{d}^d x\,(\det e)\,A^{\mu\nu\lambda}(\partial_\mu\Lambda_{\nu\lambda} + \partial_\nu\Lambda_{\lambda\mu} + \partial_\lambda\Lambda_{\mu\nu}).\end{aligned} \tag{7b}$$

In the second line of (7b) a surface term has been dropped after an integration by parts. So the theory is invariant under the extra gauges.

This gauge invariance is compatible with some additional terms in the action. For example we may add a cosmological term

$$S_1 = c_1\int \mathrm{d}^d x\,\det e. \tag{8}$$

In addition, we first note that the field strength $F_{\mu\nu\lambda\sigma}$ given by the curl of $A_{\mu\nu\lambda}$,

$$F_{\mu\nu\lambda\sigma} = (\partial_\mu A_{\nu\lambda\sigma} - \partial_\nu A_{\mu\lambda\sigma} + \partial_\lambda A_{\sigma\mu\nu} - \partial_\sigma A_{\lambda\mu\nu}), \tag{9}$$

is invariant under both the tangent Lorentz and extra gauge transformations. Thus any function of this field compatible with general coordinate invariance can be included in the action. In particular the quadratic term

$$S_2 = c_2\int \mathrm{d}^d x\,(\det e)\,F_{\mu\nu\lambda\sigma}F^{\mu\nu\lambda\sigma} \tag{10}$$

which has dimension 4 can serve as a kinetic term.

In more than 4 dimensions there are some special terms that could be added as permitted by the existence of the completely antisymmetric Levi-Civita tensor. Thus, in 7 dimensions we can consider

$$S_7 = c_7\int \mathrm{d}^7 x\,A_{\mu_1\mu_2\mu_3}F_{\mu_4\mu_5\mu_6\mu_7}\,\epsilon^{\mu_1\mu_2\mu_3\mu_4\mu_5\mu_6\mu_7}. \tag{11}$$

In 11 dimensions

$$\begin{aligned}S_{11} = c_{11}\int \mathrm{d}^{11}x\,&A_{\mu_1\mu_2\mu_3}F_{\mu_4\mu_5\mu_6\mu_7}F_{\mu_8\mu_9\mu_{10}\mu_{11}}\\ &\times\epsilon^{\mu_1\mu_2\cdots\mu_{10}\mu_{11}},\end{aligned} \tag{12}$$

etc. Note that these are gauge invariant up to a surface term, which we drop.

In addition to the usual topological quantities constructed from the curvature, new topological invariants can be obtained in special dimensions via the field strength $F_{\mu\nu\lambda\sigma}$. For example in 8 dimensions

$$S_8 = \int \mathrm{d}^8 x\,F_{\mu_1\mu_2\mu_3\mu_4}F_{\mu_5\mu_6\mu_7\mu_8}\,\epsilon^{\mu_1\mu_2\mu_3\mu_4\mu_5\mu_6\mu_7\mu_8}. \tag{13}$$

We now consider the equation of motion for the connection that follows from the action $S = S_0 + S_1 + S_2 + \ldots$ We find

$$(T_{\mu\nu}^i e_{i\lambda} - A_{\mu\nu\lambda}) = X_{\mu\nu\lambda}, \tag{14}$$

where

$$X_{\mu\nu\lambda} = -32\,c_2\,\partial^\sigma[F_{\mu\nu\lambda\sigma}\det e] + \ldots. \tag{15}$$

Note that X is a completely antisymmetric tensor, while the term in parenthesis in eq. (14) has no completely antisymmetric component. Therfore each side of eq. (14) is zero separately

$$-32\,c_2\,\partial^\sigma(F_{\mu\nu\lambda\sigma}\det e) + \ldots = 0 \tag{16}$$

and

$$T_{\mu\nu}^i e_{i\lambda} = A_{\mu\nu\lambda}. \tag{17}$$

This shows that the completely antisymmetric field $A_{\mu\nu\lambda}$ which is the only non-vanishing component of the torsion tensor $T^i_{\mu\nu} e_{i\lambda}$, can be taken as a dynamical variable whose equation of motion is given by eq. (16). Furthermore, eq. (17) determines the connection as the sum of the Riemann connection of (3) and the torsion $A_{\mu\nu\lambda}$:

$$\omega_\mu^{ij} = \overset{\circ}{\omega}{}_\mu^{ij} + A_{\mu\nu\lambda} e^{\nu i} e^{\lambda j}. \qquad (18)$$

In terms of the independent fields e^i_μ and $A_{\mu\nu\lambda}$ we may write the action in the "second order" formalism, rather than the "first order" formalism of above:

$$S_0 = -\frac{1}{4}\int \mathrm{d}^d x \det e \, \overset{\circ}{R}{}^{ij}_{\mu\nu} e_i^\mu e_j^\nu ,$$

$$\overset{\circ}{R}{}^{ij}_{\mu\nu} = \partial_\mu \overset{\circ}{\omega}{}^{ij}_\nu - \partial_\nu \overset{\circ}{\omega}{}^{ij}_\mu + [\overset{\circ}{\omega}_\mu, \overset{\circ}{\omega}_\nu]^{ij}, \qquad (19)$$

while the rest of the action S_1, S_2, etc. retain the same form. Then, the equation of motion for $A_{\mu\nu\lambda}$, i.e. $X_{\mu\nu\lambda} = 0$, can also be derived from this new form of the action.

The second order form of the action taken in 11 dimensions, with the special coefficients

$$c_1 = 0, \qquad c_2 = -\tfrac{1}{48}, \qquad c_{11} = 2/(12)^4, \qquad (20)$$

is seen to be identical to the bosonic sector of 11-dimensional supergravity [3]. Our idea was motivated by the desire to understand the geometric significance of this theory. Indeed we have succeeded in formulating it in terms of e^i_μ and ω^{ij}_μ alone.

We have shown here how one can construct a *first order* action, in any number of dimensions, in which the torsion is a completely antisymmetric tensor $A_{\mu\nu\lambda}$, and an additional gauge invariance holds with $A_{\mu\nu\lambda}$ as the gauge potential. In particular, the boson sector of 11-dimensional supergravity can be reformulated in this way. Then Englert's [4] observation that in a special solution of the equations of motion $A_{\mu\nu\lambda}$ represents the torsion of a 7-sphere, naturally follows from our formulation. Note further that, in our language, Englert's parallelizability coincides with the requirement that ω^{ij}_μ must be a pure (Lorentz) gauge in the 7-dimensional subspace.

In the absence of supersymmetry, our approach also allows a cosmological term. This implies that we may be able to find vacuum solutions which dimensionally reduce $d = 7, 11$, etc, dimensions down to 4 dimensions in such a way that the cosmological constant of the 4-dimensional space is arranged to be identically zero or very small. It would then be interesting to apply the recent ideas on the computation of the fine structure constants [5].

We are presently investigating the supersymmetric extension of our first order formalism in various dimensions. One of our aims is to analyze 11-dimensional supergravity with the purpose of getting a better understanding of its geometrical structure and hopefully a superspace formulation of the theory.

References

[1] R. Utiyama, Phys. Rev. 101 (1956) 1597.
[2] S. MacDowell and F. Mansuri, Phys. Rev. Lett. 38 (1977) 739.
[3] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.
[4] F. Englert, Phys. Lett. 119B (1982) 339.
[5] T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141;
S. Weinberg, Texas preprint (April 1983).

GAUGED $N = 8$ SUPERGRAVITY AND ITS BREAKING FROM SPONTANEOUS COMPACTIFICATION

B. BIRAN [1], F. ENGLERT
Faculté des Sciences, Université Libre de Bruxelles [2], Brussels, Belgium

B. de WIT
NIKHEF-H, Amsterdam, The Netherlands

and

H. NICOLAI
CERN, Geneva, Switzerland

Received 20 January 1983

Gauged $N = 8$ supergravity emerges from spontaneous compactification of eleven-dimensional supergravity on the riemannian S^7. This is the only non-trivial compactification scheme preserving full supersymmetry. The four-dimensional theory admits a spontaneously broken solution describable as a compactification on the parellelized S^7; supersymmetry is completely broken and the local SO(8) group is reduced to G_2.

It is well known that four-dimensional gravity, Yang–Mills interactions and matter fields may originate from a higher-dimensional theory of pure gravity [1]. This prospect for unifying matter and the fundamental interactions including gravity is also offered by supergravity, so it is a natural idea to combine these two approaches. This is even more so because supersymmetry puts restrictions on the number of possible space–time dimensions. The maximal dimension for which one can balance bosonic and fermionic degrees of freedom with highest spin two is eleven. Supergravity in eleven dimensions has been constructed some time ago [2], and it is thus possible to study spontaneous compactifications of this theory, i.e. solutions of the eleven-dimensional equations of motion for which the ground state corresponds to a product space of a four-dimensional space–time M^4 and a compact seven-dimensional space M^7

$$M^{11} \to M^4 \otimes M^7. \tag{1}$$

The equations of motion restrict M^4 to a flat Minkowski or to an Einstein space. For M^7 there are more options. The well-known hypertorus T^7 solves these equations trivially and its zero-mass sector corresponds to $N = 8$ supergravity in four dimensions [3]. However, one may envisage other possibilities, either directly motivated by phenomenological considerations [4] or inspired by more formal arguments. An example of the latter is the sphere S^7 [5,6] whose striking mathematical properties fit in a surprising way into the structure of the eleven-dimensional theory. This paper is devoted to a study of supergravity in which the seven extra dimensions are compactified to different geometries of S^7.

In general, compactification induces supersymmetry breaking. Therefore, we shall first prove that the eleven-dimensional theory permits only two supersymmetric compactifications with a four-dimensional Lorentz invariant ground state, namely T^7 or S^7, with the four-space identified respectively with Minkowski or anti-de Sitter space. We recall that eleven-dimensional supergravity is described in terms of three gauge fields: the gravitational elfbein $E_M{}^A$, the spinor gauge field of supersymmetry Ψ_M and an antisymmetric three-rank gauge field A_{MNP}. The latter has a corresponding field strength $F_{MNPQ} = 24\partial_{[M}A_{NPQ]}$. A vacuum state $|\Omega\rangle$ describing a given background can preserve supersymmetry if and only if the variation of the supercovariant Rarita–Schwinger field strength $\langle\Omega|\delta R_{AB}(\Psi)|\Omega\rangle$ vanishes for arbitrary supersymmetry transformations. This amounts to analyzing an integrability condition because

---

[1] Chercheur I.I.S.N.

[2] Supported in part by the Belgian State under the contract A.R.C. 79/83-12.

$$\delta R_{MN}(\Psi) = [\hat{D}_M, \hat{D}_N]\,\epsilon. \tag{2}$$

One finds that

$$\begin{aligned}\delta R_{AB}(\Psi) = {} & \tfrac{1}{4}R_{AB}{}^{CD}\Gamma_{CD}\,\epsilon + (\mathrm{i}\sqrt{2}/144)F^{CDEF}{}_{;[A}(\Gamma_{B]CDEF} - 8g_{B]C}\Gamma_{DEF})\,\epsilon - (144)^{-2}[F^{C_1\ldots C_4}F^{C_5\ldots C_8} \\ & \times \Gamma_{ABC_1\ldots C_8}\,\epsilon - 16\delta^{C_1}_{[A}F_{B]}{}^{C_2C_3C_4}F^{C_5\ldots C_8}\Gamma_{C_1\ldots C_8}\,\epsilon + 24(2F_A{}^{C_1\ldots C_3}F_B{}^{C_4\ldots C_6} - 3\delta^{C_1C_2}_{AB}F^{C_3C_4DE}F_{DE}{}^{C_5C_6} \\ & + 4\delta^{C_1}_{[A}F_{B]}{}^{DC_2C_3}F^{C_4\ldots C_6}{}_D)\Gamma_{C_1\ldots C_6}\,\epsilon - 384F_{ABC}{}^{C_1}F^{C_2\ldots C_4D}\Gamma_{C_1\ldots C_4}\,\epsilon + 576\,\delta^{C_1}_{[A}F_{B]}{}^{C_2DE}F^{C_3C_4}{}_{DE} \\ & \times \Gamma_{C_1\ldots C_4}\,\epsilon + 24(-36F_A{}^{CEF}F_B{}^{DEF} + \delta^{CD}_{AB}F^2 - 8\delta^C_{[A}F_{B]EFG}F^{DEFG})\Gamma_{CD}\,\epsilon].\end{aligned} \tag{3}$$

The tensors $R_{AB}{}^{CD}(E)$ and F_{ABCD} denote the supercovariant generalization of the eleven-dimensional Riemann tensor and field strength with flat indices. Here and in what follows, expectation value symbols are not exhibited.

Thus, for a fully supersymmetric compactification to exist, eq. (3) must vanish. This implies in particular that the field equations must be satisfied, because $\Gamma^{ABC}\delta R_{AB}(\Psi)$ is proportional to the bosonic field equations. In other words, the compactification must be spontaneous. Furthermore, the coefficients of independent Γ matrix combinations have to vanish independently. Since the tensors $R(E)$ and F must be invariant under four-dimensional Lorentz transformations, one can verify from eq. (3) that there is a unique solution for which the non-zero values of $R(E)$ and F are given by

$$F^{\mu\nu\rho\sigma} = fe^{-1}\epsilon^{\mu\nu\rho\sigma}, \quad R_{\mu\nu\rho\sigma}(E) = m_4^2(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}), \quad R_{mnpq}(E) = -m_7^2(g_{mp}g_{nq} - g_{mq}q_{np}) \tag{4}$$

with

$$m_7^2 = \tfrac{1}{4}m_4^2 = \tfrac{1}{18}f^2, \tag{5}$$

where four-dimensional indices are denoted by $\mu, \nu, \ldots$ and seven-dimensional ones by $m, n, \ldots$. We see that f acts as an order parameter for spontaneous compactification of eleven-dimensional supergravity to four dimensions as was pointed out in ref. [7]. The case $f = 0$ yields the well-known compactification on the torus T^7 with a minkowskian space–time. It was first realized by Duff and Pope that the case $f \neq 0$ with the additional requirement of eight unbroken supersymmetries corresponds to a spontaneous compactification on the riemannian S^7 in an anti-de Sitter space–time with radii related by eq. (5) [6].

The massless sector of the supersymmetric compactification with $f \neq 0$ is expected to correspond to $N = 8$ supergravity with local SO(8) invariance [8] in view of the fact that the group of rigid motions of S^7 is SO(8). This correspondence will be made precise below. If F_{mnpq} is non-zero then supersymmetry is broken. This happens when compactifying the seven dimensions to the parallelized S^7 [9], which, as we shall show in the last part of this letter, corresponds to a spontaneously broken solution of gauged $N = 8$ supergravity.

We now proceed to the proof that the compactification on S^7 leads to gauged $N = 8$ supergravity in four dimensions. For this purpose, we have to make a distinction between four- and seven-dimensional coordinates, denoted by x^μ and y^m, respectively. We then analyze the kinetic terms, i.e. the terms containing space–time derivatives, and redefine the fields so that there are no off-diagonal kinetic terms. For the fermionic fields, this requirement yields

$$\Psi_\mu(x,y) = \psi'_\mu(x,y) + \tfrac{1}{2}\gamma^5\gamma_\mu\Gamma^m\psi'_m(x,y), \quad \Psi_m(x,y) = \psi'_m(x,y), \tag{6}$$

where we have introduced new fields ψ'_μ and ψ'_m; this redefinition is the same as for the reduction on T^7 [3]. Note that at the linearized level there is no need to distinguish between curved and flat indices. The field redefinitions in the bosonic sector are not so straightforward. First, we re-express the metric tensor in eleven dimensions according to

$$g_{MN}(x,y) = g^{(0)}_{MN}(x,y) + h_{MN}(x,y), \tag{7}$$

where $g^{(0)}_{\mu\nu}(x,y) = g^{(0)}_{\mu\nu}(x)$ is the usual background metric in an anti-de Sitter space, and $g^{(0)}_{mn}(x,y) = g^{(0)}_{mn}(y)$ is the metric on S^7; moreover, $g^{(0)}_{\mu m}(x,y) = 0$. The necessary field redefinitions which lead to diagonal kinetic terms are then given by

$$h_{\mu\nu}(x,y) = h'_{\mu\nu}(x,y) - \tfrac{1}{2} g^{(0)}_{\mu\nu}(x)\, h'^{m}{}_{m}(x,y), \qquad h_{mn}(x,y) = h'_{mn}(x,y). \tag{8}$$

Observe that the first formula is nothing but the usual Weyl rescaling to lowest order.

It turns out that this redefinition is not sufficient to completely remove the mass-mixing between gravity and the other fields, because $F_{\mu\nu\rho\sigma}$ does not vanish in the background. To ensure complete decoupling, we must also expand the three-index auxiliary gauge field $A_{\mu\nu\rho}$

$$A_{\mu\nu\rho}(x,y) = A^{(0)}_{\mu\nu\rho}(x) + (1/3!)\, \epsilon_{\mu\nu\rho\sigma}\, t^{\sigma}(x,y) \tag{9}$$

and redefine its fluctuation according to

$$t^{\mu}{}_{;\mu}(x,y) = t'^{\mu}{}_{;\mu}(x) + \tfrac{3}{2}\sqrt{2}\, m_7 \left[h'^{\mu}{}_{\mu}(x,y) - h'^{m}{}_{m}(x,y)\right]. \tag{10}$$

In this way, gravity indeed decouples from all scalar excitations; thereby states of higher consciousness are promoted to the ultimate vacuum.

The next step in our construction is the harmonic expansion of all relevant fields on S^7. Since the four-dimensional lagrangian is recovered from the eleven-dimensional theory by integration over S^7

$$\mathcal{L}^{(4)}(x) = \int_{S^7} \mathcal{L}^{(11)}(x,y)\, g^{1/2}\, \mathrm{d}y \tag{11}$$

it is clear that the zero mass-sector of the theory is determined by suitable eigenmodes of the corresponding differential operators on S^7. We shall prove that these modes are related by supersymmetry and that the particle spectrum of $N = 8$ supergravity emerges. In general, it is very difficult to perform such an eigenmode expansion, but in this case there are two circumstances which facilitate our construction. Namely, it was proposed in ref. [6] to employ covariantly constant spinors on S^7, and this observation turns out to be the key to the construction of the relevant zero-modes on S^7, because these can be represented as products of covariantly constant spinors (this possibility has also been explored by the authors of ref. [6]). There are two kinds of covariantly constant spinors which we denote by η^I_+ and η^I_-. They provide a basis for the solution of the integrability condition (3) and satisfy the equations

$$(D_m \pm \tfrac{1}{2} m_7 \Gamma_m)\, \eta^I_\pm(y) = 0, \qquad I, \ldots = 1, \ldots, 8 \tag{12}$$

and are normalized to unity, i.e.[‡1]

$$\bar{\eta}^I_+(y)\, \eta^J_+(y) = \bar{\eta}^I_-(y)\, \eta^J_-(y) = \delta^{IJ}. \tag{13}$$

It is important here that the SO(8) indices $I, J, \ldots$ belong to the *spinorial* rather than the vectorial representation of SO(8) (otherwise, the 8×8 matrix η^I_α would provide an equivalence transformation between the two inequivalent representations). The second crucial observation is that the supersymmetry transformations allow one to deduce how the spinors should occur in the solutions. These transformations are

---

[‡1] In our notation, $\bar{\eta}^I \equiv (\eta^I)^\dagger$.

$$\delta E_M{}^A = -\tfrac{1}{2}\mathrm{i}\bar{\epsilon}\Gamma^A\Psi_M, \quad \delta A_{MNP} = \tfrac{1}{8}\sqrt{2}\,\bar{\epsilon}\Gamma_{[MN}\Psi_{P]}, \quad \delta\Psi_M = D_M(\hat{\omega})\epsilon + (\mathrm{i}\sqrt{2}/288)(\Gamma_M{}^{NPQR} - 8\delta^N_M\Gamma^{PQR})\epsilon F_{NPQR}\,. \tag{14}$$

The parameter ϵ that describes supersymmetry in four dimensions is expected to be covariantly constant on S^7, which suggests that we decompose ϵ into the spinors η^I. The associated gauge field ψ'_μ is then decomposed accordingly. The SO(8) gauge fields associated with the rigid motions of S^7 emerge from the eleven-dimensional metric in the standard way [1]. Since one anticipates gauged $N = 8$ supergravity in four dimensions, there should be 56 zero modes for the remaining spinors described by ψ'_m. This suggests that ψ'_m is decomposed into the 56 independent antisymmetric products of Killing spinors: $\psi'_m \sim \eta^{[I}\eta^J\eta^{K]}$. A comparison of these ansätze with the supersymmetry transformations [eq. (14)] then shows that the scalars and pseudoscalars which emerge from the metric tensor g_{mn} and the antisymmetric gauge field A_{mnp}, respectively, must be decomposed in terms of antisymmetrized products of four η's. The precise form of these decompositions is then established by requiring that they correspond to the massless modes. Except for the spin 0 and the spin 1/2 fields, this procedure is rather straightforward. The correct ansätze are found to be

$$h'_{\mu\nu}(x,y) = h_{\mu\nu}(x) + \dots, \quad \psi'_\mu(x,y) = \psi^I_\mu(x)\eta^I_+(y) + \dots, \quad h'_{\mu m}(x,y) = A^{IJ}_\mu(x)\bar{\eta}^I_+(y)\Gamma_m\eta^J_+(y) + \dots, \tag{15,16,17}$$

$$\psi'_m(x,y) = \chi^{IJK}(x)[\eta^{IJK}_m(y) - \tfrac{1}{9}\Gamma_m\Gamma^n\eta^{IJK}_n(y)] + \dots, \tag{18}$$

$$h'_{mn}(x,y) = A^{IJKL}(x)\{\bar{\eta}^{[I}_+(y)\Gamma_m\eta^{JKL]}_n(y) - \tfrac{1}{9}g^{(0)}_{mn}(y)\bar{\eta}^{[I}_+(y)\Gamma^P\eta^{JKL]}_P(y)] + \dots, \tag{19}$$

$$A_{mnp}(x,y) = B^{IJKL}(x)\bar{\eta}^{[I}_+(y)\Gamma_{[mn}\eta^{JKL]}_{P]}(y) + \dots, \tag{20}$$

where we have introduced the vector–spinor

$$\eta^{IJK}_m(y) \equiv \eta^{[I}_+(y)\bar{\eta}^J_+(y)\Gamma_m\eta^{K]}_+(y) \tag{21}$$

to simplify the notation. The ansätze (15) and (17) follow directly from the general theory [1]. Ansatz (16) has already been motivated; we recall that the eleven-dimensional supersymmetry transformation parameter is expressed in an analogous manner

$$\epsilon(x,y) = \epsilon^I(x)\eta^I_+(y) + \dots \tag{22}$$

which makes the eight supersymmetries of the theory manifest. The supersymmetry transformations which correspond to the higher modes in the expansion of $\epsilon(x,y)$ are not symmetries of the truncated four-dimensional action, since they connect the different sectors that arise in the harmonic expansion on S^7. The full set of supersymmetry transformations will generate a generalized super-Kac–Moody algebra (with the unit circle S^1 replaced by S^7), which is obtained by projecting the commutator of the eleven-dimensional theory onto the various eigenmodes on S^7. In this way, one recognizes immediately that there exists a one-to-one correspondence between the fermionic generators of this algebra and the eigenmodes of the Dirac operator on S^7 which appear in eq. (22). Note that this realization of an infinite dimensional superalgebra is quite different from the ones that have been previously considered [10].

From eq. (17) we see that there are indeed 56 spin 1/2 states. The 35 scalars contained in h'_{mn} correspond to self-dual antisymmetric tensors; their self-duality follows from the identities of ref. [3]. The 35 pseudoscalars B^{IJKL} have the opposite duality phase. Note that h'_{mn} is *not* a Killing tensor. We remark that there exists an alternative ansatz for the pseudoscalars in terms of the symmetric trace-free 35-representation of SO(8), namely [6]

$$A_{mnp}(x,y) = B^{IJ}(x)\bar{\eta}^I_-(y)\Gamma_{mnp}\eta^J_-(y), \tag{23}$$

where the spinors of negative parity are used. Both eqs. (20) and (23) obey the same masslessness condition

$$D_m A_{npq} = \tfrac{1}{6}m_7\,\epsilon_{mnpqrst}A^{rst} \tag{24}$$

and are therefore linearly dependent. The ansatz (23) will be useful for the discussion of the broken case.

Observe that the trace terms in eqs. (18) and (19), which are necessary to ensure masslessness, are also needed to get the correct four-dimensional transformation laws. For example, from eqs. (14) and (18), we find after a Fierz rearrangement of the S^7 spinors that

$$\delta A_{mnp}(x,y) = \tfrac{1}{8}\sqrt{2}\,\bar{\epsilon}(x,y)\Gamma_{[mn}\psi'_{p]}(x,y) = \tfrac{1}{8}\sqrt{2}\,\bar{\epsilon}^{I}(x)\chi^{JKL}(x)(\bar{\eta}^{I}_{+}\Gamma_{[mn}\eta^{JKL}_{p]} - \tfrac{1}{9}\bar{\eta}^{I}_{+}\Gamma_{mnp}\Gamma^{q}\eta^{JKL]}_{q}) + \dots$$
$$= \tfrac{1}{12}\sqrt{2}\,\bar{\epsilon}^{[I}(x)\chi^{JKL]}(x)\bar{\eta}^{I}_{+}\Gamma_{[mn}\eta^{JKL}_{p]} + \dots, \tag{25}$$

which is the desired result.

One can now show that the ansätze (15)–(20) describe the massless modes. We emphasize that masslessness in anti-de Sitter space does not mean that mass terms are completely absent. In a gravitational background the wave equation of spinless fields contains a term $R/6$, where R is the four-dimensional curvature scalar; moreover, there is an apparent mass term for the spin 3/2 fields. Because the massless modes are related by supersymmetry and constitute the full spectrum of $N = 8$ supergravity, we conclude that the lagrangian of $N = 8$ supergravity in four dimensions is obtained to second and therefore to all orders after integration over S^7.

In the last part of this letter we prove that gauged $N = 8$ supergravity can be spontaneously broken by a mechanism which has geometrical significance in eleven dimensions. To this effect we express $A_{mnp}(x,y)$ in the η^{-} basis according to eq. (23), where the $B^{IJ}(x)$ are linear combinations of the $B^{IJKL}(x)$. The quantity

$$S_{mnp}(y) \equiv m_7\bar{\psi}\Gamma_{mnp}\psi, \quad \bar{\psi}\psi = 1, \tag{26}$$

where ψ is an arbitrary linear combination of the η^{I}_{-}, obviously solves the duality equation (24). We shall prove that this seven-parameter submanifold of solutions of eq. (24) consists of the (left) torsions that parallelize S^7; in fact, eq. (26) solves the classical equations of motion provided that we choose [9]

$$m_7^2 = \tfrac{3}{10}m_4^2 = \tfrac{1}{8}f^2, \quad F_{mnpq} = \pm(\sqrt{2}/m_7)D_m S_{npq} = \pm\tfrac{1}{6}\sqrt{2}\,\epsilon_{mnpqrst}S^{rst}. \tag{27}$$

The field strength F_{mnpq} does not vanish and thus supersymmetry is broken. Note that eq. (26) can be interpreted as an expectation value of the pseudoscalar fields B^{IJKL}, because the y dependence is still given by eq. (20) (this interpretation was first suggested in ref. [11]). Exploiting the fact that eqs. (20) and (23) are linearly dependent, it is possible to compute the vacuum expectation value of the pseudoscalar fields explicitly. For suitably chosen ψ, this expectation value becomes proportional to

$$(\Gamma_{mnp})_{88}\Gamma^{mn}_{[IJ}\Gamma^{p}{}_{KL]}\,. \tag{28}$$

Therefore, this classical solution describes a spontaneously broken realization of gauged $N = 8$ supergravity in four dimensions, whose origin may be traced to a "spontaneously induced parallelism" on S^7. Actually, given the solution (28), one can now proceed to analyze the symmetry breaking purely within the four-dimensional context and to verify that eq. (28) indeed leads to a stationary point of the four-dimensional scalar field potential [8]. It is also straightforward to calculate the expectation value of the A_1 and A_2 tensors of ref. [8], and thus the various mass matrices of the four-dimensional theory from eq. (28).

The tensors S_{mnp} satisfy the following identities:

$$S_{mnt}S^{pqt} = -2m_7^2\delta^{pq}_{mn} - \tfrac{1}{6}m_7\,\epsilon_{mn}{}^{pqrst}S_{rst}, \quad S^{[mnp}S^{q]\,rs} = -\tfrac{1}{4}m_7\,\epsilon^{mnpq[r}{}_{ab}S^{s]\,ab}. \tag{29,30}$$

From eq. (29), one shows that

$$S_{mnp;q} = S_{t[mn}S_{p]\,q}{}^{t}, \quad S_{tmn}S^{t}{}_{pq} = S_{mnp;q} - m_7^2(g_{mp}g_{nq} - g_{mq}g_{np}), \tag{31}$$

which are indeed the necessary and sufficient conditions for S_{mnp} to be a parallelizing torsion [9,12].

We now turn to the characterization of the symmetry breaking. The little group of ψ at a given point y is known to be G_2 [13]. Hence, this group leaves the torsion invariant at this point and induces global motions on S^7 which map every torsion to an equivalent one. Therefore, the group of rigid motions on S^7 is reduced from

SO(8) to G_2, and the gauge symmetry is reduced accordingly.

Finally, we discuss how supersymmetry is broken by the solution (26). Therefore, we return to the integrability condition, eq. (3), and insert the value of F_{mnpq} associated with eq. (26). We then evaluate the resulting expression by using the identities (29) and (30). From the fact that the field equations are satisfied, we infer that it is sufficient to analyse eq. (3) for seven-dimensional indices only. After contraction with Γ^{mn}, we find that eq. (3) becomes

$$\Gamma^{mn}\delta R_{mn}(\Psi) = -\tfrac{1}{3}(14m_7^2 + 10m_7\Gamma^{mnp}S_{mnp})\epsilon . \tag{32}$$

The eigenvectors of the matrix $(\Gamma^{mnp}S_{mnp})$ are

$$(\Gamma^{mnp}S_{mnp})\psi = -42m_7\psi , \quad (\Gamma^{mnp}S_{mnp})\Gamma_q\psi = 6m_7\Gamma_q\psi , \tag{33}$$

where the spinors ψ and $\Gamma_m\psi$ span the one-dimensional and seven-dimensional subspaces that are left invariant by the action of G_2. Clearly, eq. (32) has no zero eigenvalues, so that all eight supersymmetries are spontaneously broken. The eigenvalues of the matrix in eq. (32) characterize the scales at which the supersymmetries are broken. As a consequence of eq. (32), seven supersymmetries are broken at the same scale. Altogether, there are thus two supersymmetry-breaking scales which are expressed in terms of the S^7 radius through eqs. (32) and (33).

We conclude with a comment which may be relevant for the phenomenological interpretation of $N = 8$ supergravity. So far, it has been generally assumed that the higher modes of Kaluza–Klein theories may be neglected in the analysis of the low-energy sector. While this is undoubtedly true for the unbroken theory, it is not necessarily true for the broken theory. Since the whole mass spectrum of the theory is shifted by units of the order of the inverse S^7 radius by the symmetry breaking, one cannot a priori rule out the possibility that some of the previously massive states become massless. Such a "level crossing" would obviously affect the phenomenology at low energies, and therefore this possibility deserves further investigation.

A detailed account of the results reported here will be published elsewhere.

We are grateful to M.J. Duff for his collaboration at early stages of this work.

[1] Th. Kaluza, Sitzungsber. Preus. Akad. Wis. Berlin K1 (1921) 966;
O. Klein, Z. Phys. 37 (1926) 895;
B. de Witt, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965);
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Rep. Math. Phys. 1 (1970) 29;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316.

[2] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.

[3] E. Cremmer and B. Julia, Phys. Lett. 80B (1978) 48; Nucl. Phys. B159 (1979) 141.

[4] E. Witten, Nucl. Phys. B186 (1981) 412.

[5] W. Nahm, CERN preprint TH. 2483 (1978), unpublished.

[6] M.J. Duff and C.N. Pope, to be published, reported by M.J. Duff, in: Supergravity 81, eds. S. Ferrara and J.G. Taylor (Cambridge U., London, 1982);
M.J. Duff and D.J. Toms, in: Unification of the fundamental interactions II, eds. S. Ferrara and J. Ellis (Plenum, New York, 1982); CERN preprint TH. 3259 (1982).

[7] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233.

[8] B. de Wit and H. Nicolai, Phys. Lett. 108B (1981) 285; Nucl. Phys. B208 (1982) 323.

[9] F. Englert, Phys. Lett. 119B (1982) 339.

[10] M. Günaydin, in: Unification of the fundamental interactions II, eds. S. Ferrara and J. Ellis (Plenum, 1982); CERN preprint TH. 3222 (1982);
B. Julia, Kac–Moody symmetry of gravitation and supergravity theories, LPTENS preprint 82/22 (1982).

[11] M.J. Duff, CERN preprint TH. 3451 (1982).

[12] E. Cartan and J. Schouten, Proc. Kon. Akad. Wet. Amsterdam 29 (1926) 933.

[13] R.E. Behrends, J. Dreitlein, C. Fronsdal and W. Lee, Rev. Mod. Phys. 34 (1962) 1.

THE MASS SPECTRUM OF SUPERGRAVITY ON THE ROUND SEVEN-SPHERE

A. CASHER[1] and F. ENGLERT
Université Libre de Bruxelles

H. NICOLAI
CERN, Geneva, Switzerland

M. ROOMAN
Université Libre de Bruxelles

Received 7 March 1984

We derive the mass spectrum of supergravity compactified on the round seven-sphere. The final result may be arrived at either by employing harmonic expansions on S^7 or by using properties of Osp(8, 4).

1. Introduction

Simple supergravity in eleven space-time dimensions [1] naturally permits spontaneous compactification [2] to four space-time dimensions because of the presence of a three-index gauge-field in that theory [3]. This means that the field equations of $d = 11$ supergravity have solutions by which $d = 11$ space-time $\mathcal{M}_{11}$, which is locally parametrized by coordinates z^M, spontaneously decomposes into a product of a four-dimensional space-time $\mathcal{M}_4$ and an "internal" manifold $\mathcal{M}_7$, which are locally parametrized by x^μ ($\mu = 0, 1, 2, 3$) and y^m ($m = 5, 6, \dots, 11$), respectively. There are only two fully supersymmetric compactifications with $n = 8$ supersymmetry [4], namely one with $\mathcal{M}_4 =$ Minkowski space and $\mathcal{M}_7 = T^7$ [5] and the other with $\mathcal{M}_4 = (\mathrm{AdS})_4$ and $\mathcal{M}_7 = S^7$ [6]. In the truncation where the y-dependence is discarded, the first leads to ungauged $N = 8$ supergravity [5]; the massive modes correspond to the Fourier coefficients on T^7. In the second case, as well as in more complicated cases, one must expand the fields of $d = 11$ supergravity, which we collectively denote by $\phi(x, y)$, in terms of a suitable complete set of eigenfunctions $Y^{(n)}(y)$ of the relevant mass operator according to

$$\phi(x, y) = \sum_n \phi^{(n)}(x)\, Y^{(n)}(y)\,. \tag{1.1}$$

It has been demonstrated in refs. [4, 6] that, for $\mathcal{M}_7 = S^7$, there indeed occurs a massless $N = 8$ supermultiplet in the expansion (1.1). Since the isometry group of S^7 is SO(8), the effective $d = 4$ theory which is obtained in the truncation where the higher modes in the expansion (1.1) are discarded is presumably equivalent to gauged $N=8$ supergravity [7], at least if non-linear modifications are properly taken into account. According to the conventional lore [8], the excited modes in the expansion (1.1) describe particles with masses of the order of the Planck mass and multiples thereof and therefore should play no role in the analysis of the low-energy behaviour of the theory. However, they are expected to play an important role at the quantum level [9]. Moreover, it is doubtful that the conventional philosophy is entirely correct since the natural unit of mass which arises in the compactification is of the order of the inverse size of the internal manifold $\mathcal{M}_7$; the latter being a dynamical parameter in the quantized theory, there is a priori no reason to take it to be of the order of the Planck mass [10]. It is therefore of interest to determine the full mass spectrum for the S^7 compactification and other cases. In this paper, we present the calculation of the mass spectrum on $\mathcal{M}_7 = S^7$ in detail. For the bosonic modes, the results have already been given in ref. [11] while the complete results have been reported in ref. [10]*. For manifolds other than S^7, so far only partial results have been obtained: the zero modes on the squashed S^7 [14], which constitute the massless supermultiplets, have been derived in ref. [15]; the massive spin-2 and spin-$\frac{3}{2}$ modes on the squashed S^7 have been determined in ref. [16] while the massive spin-1 modes were analyzed in ref. [17]; finally, the fermionic spectrum on M^{pqr} spaces has been completely determined in ref. [13]. All calculations so far have been based on harmonic expansions but we believe that the group theoretical methods described in sect. 3 of this paper will also be useful for other supersymmetric compactifications: if there is a residual N-extended supersymmetry, the massive states must belong to multiplets of Osp $(N, 4)$.

---

[1] On leave of absence from Tel-Aviv University.

We now briefly summarize our conventions and notations. Supergravity in eleven dimensions [1] is based on the following multiplet of fields: an elfbein $e_M{}^A$ (flat indices are labelled by the first letters of the alphabet), a 32-component Majorana vector spinor ψ_M and an antisymmetric three-index tensor A_{MNP} which is subject to the abelian gauge transformations**.

$$\delta A_{MNP} = D_{[M}\Lambda_{NP]} \, . \tag{1.2}$$

Defining $F_{MNPQ} \equiv 24\partial_{[M}A_{NPQ]}$ and $\tilde{\Gamma}_{M_1\cdots M_k} \equiv \tilde{\Gamma}_{[M_1}\cdots\tilde{\Gamma}_{M_k]}$, the invariant action of $d=11$ supergravity reads

$$\begin{aligned} S = \int \mathrm{d}^{11}z\, \mathrm{e}\Big\{ & -\tfrac{1}{2}R - \frac{1}{48}F_{MNPQ}F^{MNPQ} \\ & -\tfrac{1}{2}i\bar{\psi}_M\tilde{\Gamma}^{MNP}\psi_{P;N} + \frac{4\sqrt{2}}{(4!)^3}\,\eta^{M_1\cdots M_{11}}F_{M_1\cdots M_4}F_{M_5\cdots M_8}A_{M_9\cdots M_{11}} \\ & + \frac{3\sqrt{2}}{(4!)^2}(\bar{\psi}_M\tilde{\Gamma}^{MNPQRS}\psi_N + 12\bar{\psi}^P\Gamma^{QR}\psi^S)F_{PQRS}\Big\} \end{aligned} \tag{1.3}$$

---

* The results of refs. [10, 11] have meanwhile been independently obtained in ref. [12] and, for the fermionic modes, in ref. [13].

** Covariant derivatives are denoted by $D_M\phi \equiv \phi_{\,M}$.

up to higher order fermionic terms. Eq. (1.3) is invariant under the local supersymmetry transformations

$$\delta e_M{}^A = -\tfrac{1}{2} i \bar{\varepsilon} \tilde{\Gamma}^A \psi_M, \quad \delta A_{MNP} = \frac{\sqrt{2}}{8} \bar{\varepsilon} \tilde{\Gamma}_{[MN}\psi_{P]}$$

$$\delta\psi_M = D_M\varepsilon + \frac{2\sqrt{2}}{(4!)^2} i(\tilde{\Gamma}_M{}^{NPQR} - 8\delta_M^N \tilde{\Gamma}^{PQR})\varepsilon F_{NPQR} . \tag{1.4}$$

Again, we have omitted higher-order fermionic terms in (1.4). In the absence of fermion condensates, the classical equations of motion, which follow from the action (1.3), are

$$R_{MN} - \tfrac{1}{2} g_{MN} R = -\tfrac{1}{48}\{8 F_{MPQR} F_N{}^{PQR} - g_{MN} F^2\} , \tag{1.5}$$

$$F^{MNPQ}{}_{;M} = -\frac{\sqrt{2}}{2\cdot(4!)^2} \eta^{NPQM_1\cdots M_8} F_{M_1\cdots M_4} F_{M_5\cdots M_8} , \tag{1.6}$$

and, up to a supersymmetry transformation, we have put

$$\psi_M = 0. \tag{1.7}$$

The solutions describing spontaneous compactification on $\mathcal{M}_4 \times \mathcal{M}_7$ are characterized by $g_{\mu m} = F_{\mu mnp} = F_{\mu\nu np} = F_{\mu\nu\rho p} = 0$. If we impose in addition

$$F_{mnpq} = 0 , \tag{1.8}$$

we get from (1.5) and (1.6)

$$\begin{aligned} F_{\mu\nu\rho\sigma} &= f\eta_{\mu\nu\rho\sigma}, \quad f = \text{constant} , \\ R_{mn} &= -6m_7^2 g_{mn} , \\ R_{\mu\nu} &= 12 m_7^2 g_{\mu\nu} , \quad m_7^2 = \tfrac{1}{18} f^2 . \end{aligned} \tag{1.9}$$

These solutions were obtained by Freund and Rubin [3]. For the purposes of this paper, we will further specialize to $\mathcal{M}_4 = (\mathrm{AdS})_4$ and $\mathcal{M}_7 = \mathrm{S}^7$ [6]. In that case the Riemann curvature tensors of $\mathcal{M}_4$ and $\mathcal{M}_7$ are given by

$$\begin{aligned} R_{\mu\nu\rho\sigma} &= 4m_7^2(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}) , \\ R_{mnpq} &= -m_7^2(g_{mp}g_{nq} - g_{mq}g_{np}) . \end{aligned} \tag{1.10}$$

Finally, we note that the $\tilde{\Gamma}$-matrices in $d = 11$ can be expressed in terms of $d = 4$ γ-matrices and eight-dimensional Γ-matrices, which generate the Clifford algebra in seven dimensions, according to ref. [5]:

$$\tilde{\Gamma}_\mu = \gamma_\mu \otimes \mathbb{1} , \quad \tilde{\Gamma}_m = \gamma^5 \otimes \Gamma_m . \tag{1.11}$$

2. The mass spectrum

The bare mass spectrum of the four-dimensional theory can be obtained by varying the fields g_{MN}, A_{MNP} and ψ_M around their background values. If there are no fermionic condensates, the bosonic field equations are given by

$$\delta(R_{MN}-\tfrac{1}{2}g_{MN}R)=-\tfrac{1}{48}\delta\{8F_{MPQR}F_N{}^{PQR}-g_{MN}F^2\}\,, \tag{2.1}$$

$$\delta(F^{MNPQ}{}_{;M})=-\frac{\sqrt{2}}{2(4!)^2}\,\delta\{\eta^{NPQM_1\cdots M_8}F_{M_1\cdots M_4}F_{M_5\cdots M_8}\}\,, \tag{2.2}$$

and are linear in δg_{MN} and δA_{MNP}. Since the fermionic fields vanish in the S^7 background, the fermionic field equation is straightforwardly obtained from the action (1.3); it reads:

$$-i\tilde{\Gamma}^{MNP}\psi_{P;N}+\tfrac{1}{96}\sqrt{2}\{\tilde{\Gamma}^{MNPQRS}+12g^{MP}g^{NQ}\tilde{\Gamma}^{RS}\}\mathring{F}_{PQRS}\psi_N=0\,, \tag{2.3}$$

where the superscript 0 labels the background value. In eqs. (2.1)–(2.3), one expands the components of ψ_M, δg_{MN} and δA_{MNP} in complete sets of eigenfunctions $Y^{(n)}(y)$ of suitable differential operators on S^7 according to (1.1). The masses of the four-dimensional fields associated with the functions $\phi^{(n)}(x)$ are then obtained in terms of the eigenvalues belonging to the eigenfunctions $Y^{(n)}(y)$. However, the actual calculation reveals a mixing between the various fields of the theory and one has in fact to diagonalize a mass matrix.

The diagonalization can be performed by redefining the fields, and such a method was used to obtain the fields describing the zero-mass supermultiplet, namely one graviton, eight gravitinos, 28 SO(8) gauge fields, 56 spin-$\frac{1}{2}$ fields, 35 scalars and 35 pseudoscalars [4, 6]. For instance, the decoupling between the spin-$\frac{3}{2}$ and spin-$\frac{1}{2}$ fields was achieved by the field redefinitions

$$\psi_\mu=\psi'_\mu+\tfrac{1}{2}\gamma^5\gamma_\mu\Gamma^m\psi'_m\,, \tag{2.4}$$

$$\psi_m=\psi'_m\,. \tag{2.5}$$

However, this procedure has the disadvantage that it only works in the truncation where massive modes are discarded. If the latter are retained, it turns out that the coupling between massless and massive modes persists. One may try to extend the field redefinitions (2.4) and (2.5) so as to achieve complete decoupling, but in that case mass-dependent and hence non-local modifications are required.

Another procedure is suggested by the very fact that the linearized equations (2.1), (2.2) and (2.3) contain spurious modes associated with the various gauge invariances of the eleven-dimensional theory. These are the modes

$$h^{(G)}_{MN}=2\xi_{(M;N)}\,, \tag{2.6}$$

$$a^{(G)}_{MNP}=\Lambda_{[MN;P]}\,,\quad \Lambda^{MN}{}_{;M}=0\,, \tag{2.7}$$

$$\psi^{(G)}_M=\mathring{\mathscr{D}}_M\varepsilon=\varepsilon_{;M}+\frac{2\sqrt{2}}{(4!)^2}\,i(\tilde{\Gamma}_M{}^{NPQR}-8\delta^N_M\Gamma^{PQR})\mathring{F}_{NPQR}\varepsilon\,. \tag{2.8}$$

Here $h_{MN}\equiv-\delta g_{MN}$ ($h^{MN}\equiv\delta g^{MN}$), a_{MNP} and the superscript (G) labels the eleven coordinate, the 45 Maxwell and the 32 supersymmetric gauge modes. These gauge modes introduce an arbitrariness in the formulation of the gauge invariant physical states in terms of fields. Therefore, by choosing 11+45+32 convenient gauge conditions we cannot only eliminate spurious modes but also diagonalize the mass matrix of the physical states at the outset. Convenient gauge choices are:

$$h^{m\mu}{}_{;m} = h^{\nu\mu}{}_{;\nu} = 0 \quad \text{(10 conditions)}\,, \tag{2.9}$$

$$h^{M}{}_{M} = 0 \quad \text{(one condition)}\,, \tag{2.10}$$

$$a^{mnp}{}_{;p} = a^{m\mu p}{}_{;p} = a^{\mu\nu p}{}_{;p} = 0 \quad \text{(45 conditions)}\,, \tag{2.11}$$

$$\tilde{\Gamma}^{M}\psi_{M} = 0 \quad \text{(32 conditions)}\,. \tag{2.12}$$

The bosonic spectrum was computed in ref. [11], but for the sake of completeness we list here the (mass)^2 values found there. These (mass)^2 are defined by the following differential operators

$$\text{spin-0} \quad \eta^{;\mu}{}_{;\mu} - 8m_7^2\eta = -m_0^2\eta\,, \tag{2.13}$$

$$\text{spin-1} \quad (\eta_{\mu;\nu} - \eta_{\nu;\mu})^{;\nu} = -m_1^2\eta_{\mu}\,, \tag{2.14}$$

$$\text{spin-2} \quad \eta_{(\mu\nu);\rho}{}^{;\rho} - 16m_7^2\eta_{(\mu\nu)} = -m_2^2\eta_{(\mu\nu)}\,, \tag{2.15}$$

$$(\eta^{\mu}{}_{\mu} = \eta^{\mu\nu}{}_{;\nu} = 0)\,.$$

In this way, the spin-0 and spin-1 fields of the "massless" supermultiplet have indeed zero mass while the (mass)^2 of the (non-conformal) graviton has been assigned the value $8m_7^2$. In units of m_7^2, the results are:

$$\begin{aligned}
m_2^2 &= (k+3)^2 - 1\,, & k &\geq 0\,,\\
m_{1^+}^2 &= (k+3)^2 - 1\,, & k &\geq 1\,,\\
(m_{1^-}^{(1)})^2 &= k^2 - 1\,, & k &\geq 1\,,\\
(m_{1^-}^{(2)})^2 &= (k+6)^2 - 1\,, & k &\geq 1\,,\\
(m_{0^-}^{(1)})^2 &= k^2 - 1\,, & k &\geq 1\,,\\
(m_{0^-}^{(2)})^2 &= (k+6)^2 - 1\,, & k &\geq 1\,,\\
(m_{0^+}^{(1)})^2 &= (k-3)^2 - 1\,, & k &\geq 2\,,\\
(m_{0^+}^{(2)})^2 &= (k+9)^2 - 1\,, & k &\geq 0\,,\\
(m_{0^+}^{(3)})^2 &= (k-3)^2 - 1\,, & k &\geq 2\,,
\end{aligned} \tag{2.16}$$

where the superscripts label different towers with the same spin-parity assignments. The massless supermultiplet is given by the lowest value of k in the towers m_2, $m_{1^-}^{(1)}$, $m_0^{(1)}$ and $m_{0^+}^{(1)}$. In these equations the integer k labels the relevant tensor harmonics on S^7 in the expansion (1.1). Note the appearance of an additional zero-mass state for $k = 4$ and a multiplet with $m^2 = -1$ for $d = 3$ in the scalar tower $m_{0^+}^{(1)}$.

The tensor and pseudo-vector potential eigenfunctions are respectively the traceless transverse $h_{\mu\nu}$ and the transverse $A_{\mu mn}$. The two vector towers are characterized by dual field strength eigenfunctions $V^{(1)}_{m\mu\nu}$ and $V^{(2)}_{m\mu\nu}$ expressed by

$$V^{(1)}_{m\mu\nu} = \mathscr{H}_{m\mu\nu} + 36(k+5)m_7^2 a_{m\mu\nu} \quad (k \geqslant 1)\,,$$
$$V^{(2)}_{m\mu\nu} = \mathscr{H}_{m\mu\nu} - 36(k+1)m_7^2 a_{m\mu\nu} \quad (k \geqslant 1)\,, \tag{2.17}$$

where

$$\mathscr{H}_{m\mu\nu} = 3\sqrt{2}\, m_7 \eta_{\mu\nu\rho\sigma} h_m{}^{\rho;\sigma}\,. \tag{2.18}$$

The two pseudoscalar towers are described by the eigenfunctions $a^{(1)}_{mnp}$ and $a^{(2)}_{mnp}$ with opposite duality phase, namely

$$6m_7(k+3)a^{(1)}_{mnp} = \eta_{mnpqrst} a^{(1)qrs;t} \quad (k \geqslant 1)\,,$$
$$6m_7(k+3)a^{(2)}_{mnp} = -\eta_{mnpqrst} a^{(2)qrs;t} \quad (k \geqslant 1)\,. \tag{2.19}$$

The fields $s^{(1)}$ and $s^{(2)}$ corresponding to the first two scalar towers are:

$$s^{(1)} = h^{mn}{}_{;m;n} - km_7^2 h^m{}_m \quad (k \geqslant 2)\,,$$
$$s^{(2)} = h^{mn}{}_{;m;n} + (k+6)m_7^2 h^m{}_m \quad (k \geqslant 0)\,, \tag{2.20}$$

while the last scalar tower is described by the traceless transverse part of h_{mn}. (The tower $s^{(1)}$ starts only at $k=2$ because it is not possible to construct a tensor h_{mn} on S^7 which leads to a non-vanishing $s^{(1)}$ when $k=0$ or 1.)

To compute the fermionic mass spectrum, we substitute the Freund-Rubin expectation value for F_{MNPQ} into (2.3) and write out four- and seven-dimensional indices explicitly*:

$$\gamma^{\mu\nu\rho}\psi_{\rho;\nu} + \gamma^5\gamma^{\mu\nu}\Gamma^m\psi_{m;\nu} - \gamma^5\gamma^{\mu\nu}\Gamma^m\psi_{\nu;m} + \gamma^\mu\Gamma^{mn}\psi_{m;n} + \tfrac{3}{2}m_7\gamma^5\gamma^{\mu\nu}\psi_\nu = 0\,, \tag{2.21}$$

$$\gamma^5\gamma^{\mu\nu}\Gamma^m\psi_{\nu;\mu} + \gamma^\mu\Gamma^{mn}\psi_{\mu;n} - \gamma^\mu\Gamma^{mn}\psi_{n;\mu} + \gamma^5\Gamma^{mnp}\psi_{p;n} - \tfrac{3}{2}m_7\gamma^5\Gamma^{mn}\psi_n = 0\,. \tag{2.22}$$

Clearly, eqs. (2.21) and (2.22) mix four-dimensional spin-$\frac{3}{2}$ and spin-$\frac{1}{2}$ fields. To unmix these equations, we make use of the gauge condition (2.12). The fermionic spurious modes are of the form $\mathring{\mathscr{D}}_{\mathcal{M}}\chi$, see Eq. (2.8). In the four- and seven-dimensional subspaces, (2.8) becomes, respectively,

$$\mathring{\mathscr{D}}_\mu\chi = \chi_{;\mu} - m_7\gamma^5\gamma_\mu\chi\,, \tag{2.23}$$
$$\mathring{\mathscr{D}}_m\chi = \chi_{;m} - \tfrac{1}{2}m_7\Gamma_m\chi\,. \tag{2.24}$$

The gauge condition (2.12) can be re-expressed in the form

$$\gamma^\mu\psi_\mu + \gamma^5\Gamma^m\psi_m = 0\,. \tag{2.25}$$

---

* Covariant derivatives on vector-spinors are fully covariantized, i.e. covariant with respect to local Lorentz as well as general coordinate transformations.

Evidently, this gauge mixes four- and seven-dimensional subspaces. Combining (2.25) with the equation of motion (2.3), one deduces an equivalent condition

$$\begin{aligned}\psi^M{}_{;M} &= \psi^\mu{}_{;\mu} + \psi^m{}_{;m}\\ &= \tfrac{1}{2}\gamma^5\gamma^\mu\psi_\mu = -\tfrac{1}{2}\Gamma^m\psi_m\,. \end{aligned} \tag{2.26}$$

Using (2.23) and (2.24), we may therefore re-express the gauge condition (2.12) as

$$\mathring{\mathscr{D}}^M\psi_M = \mathring{\mathscr{D}}^\mu\psi_\mu + \mathring{\mathscr{D}}^m\psi_m = 0\,. \tag{2.27}$$

Inserting these conditions into the equations of motion (2.21) and (2.22), we obtain

$$-\gamma^\nu\psi^\mu{}_{;\nu} - 2m_7\gamma^5\gamma^\mu\gamma^\nu\psi_\nu = -\tfrac{3}{2}m_7\gamma^5\psi^\mu + \gamma^5\Gamma^m\psi^\mu{}_{;m}\,, \tag{2.28}$$

$$-\gamma^5\gamma^\mu\psi^m{}_{;\mu} = \Gamma^n\psi^m{}_{;n} - m_7(\Gamma^m\Gamma^n\psi_n - \tfrac{3}{2}\psi^m)\,. \tag{2.29}$$

Consequently, the equations for the four-dimensional spin-$\frac{1}{2}$ and gravitino fields have been completely decoupled by the gauge choice (2.12). Inspection now shows that, up to an additive constant, the spin-$\frac{3}{2}$ mass matrix is given by the eigenvalues of the Dirac operator on S^7. Similarly, the spin-$\frac{1}{2}$ mass matrix is given by the eigenvalues of the operator on the right-hand side of (2.19). This operator cannot be directly identified with the Rarita–Schwinger operator on S^7 but may be interpreted as the Rarita–Schwinger operator plus a "de Sitter" mass term on S^7 in a special gauge.

To solve the eigenvalue equations (2.28) and (2.29), we make use of the spherical scalar and vector harmonics as well as covariantly constant spinors on S^7 [18]. The spherical harmonics $Y(y)$ and $Y_m(y)$ obey the equations

$$Y^{;p}{}_{;p} = k(k+6)m_7^2 Y \qquad (k \geqslant 0)\,, \tag{2.30}$$

$$Y_m{}^{;p}{}_{;p} = [k(k+6)-1]m_7^2 Y_m \;(k \geqslant 1)\,, \tag{2.31}$$

where Y_m is transverse, i.e. $Y^m{}_{;m} = 0$. The eigenmodes of the Dirac operator on S^7 are easily found by making the ansatz

$$\varphi(y) = \alpha m_7 Y(y)\eta(y) + \beta\Gamma^m Y_{,m}(y)\eta(y)\,, \tag{2.32}$$

where $\eta(y)$ is the covariantly constant spinor on S^7. After a little calculation, one finds that, for each $k \geqslant 1$, there are two eigenvalues whereas there is only one for $k = 0$, namely*,

$$\begin{aligned}&\lambda = \tfrac{7}{2} \quad \text{for} \quad k = 0\,,\\ &\lambda = k + \tfrac{7}{2}\,, \quad -k - \tfrac{5}{2} \quad \text{for} \quad k \geqslant 1\,, \end{aligned} \tag{2.33}$$

with

$$\alpha\beta^{-1} = \lambda + \tfrac{5}{2} \quad (\lambda \neq -\tfrac{5}{2})\,. \tag{2.34}$$

---

* These values have also been quoted in ref. [19].

The spin-$\frac{1}{2}$ mass matrix is obtained in an analogous fashion, although the calculation is considerably more tedious. To determine the mass spectrum, it follows from (2.29) that one has to solve the eigenvalue equation

$$\Gamma^n \psi^m{}_{;n} - m_7 \Gamma^m \Gamma^n \psi_n = \lambda m_7 \psi^m \,. \tag{2.35}$$

It is here that we need the vector spherical harmonics (2.31); the correct ansatz is

$$\begin{aligned}\psi_m = {}& \alpha m_7^2 Y_m \eta + \beta m_7 Y_{m;n} \Gamma^n \eta + \gamma m_7^2 Y^n \Gamma_{mn} \eta \\ & + \delta m_7 Y^{p;n} \Gamma_{mnp} \eta + \varepsilon m_7 Y_{n;m} \Gamma^n \eta + \zeta Y^{p;n}{}_{;m} \Gamma_{np} \eta \,.\end{aligned} \tag{2.36}$$

We first note that, upon applying the operator on the left-hand side of (2.35) to the ansatz (2.36), only terms of the type already present in (2.36) are produced; on the other hand, all six terms in (2.36) and, in particular, the last term containing *two* derivatives are needed. To prove this, one has to make repeated use of (2.31) and the commutator relation for two covariant derivatives in the form

$$[D_n, D_p] Y_m = -2 m_7^2 g_{m[n} Y_{p]} \tag{2.37}$$

and analogous ones for higher-order tensors.

Substituting the ansatz (2.36) into (2.35), one gets a linear equation for the six coefficients α, β, γ, δ, ε and ζ. The mass eigenvalues λ are then given by the eigenvalues of the corresponding 6×6 matrix. To facilitate the computation, it proves advantageous to first separate off the spurious eigenmodes which are of the form (2.24). Applying the operator $\mathring{\mathscr{D}}_m$ to the eigenvalue equation (2.35), we obtain

$$\Gamma^m (\mathring{\mathscr{D}}^n \psi_n)_{;m} = (\lambda + 1) m_7 \mathring{\mathscr{D}}^n \psi_n \,. \tag{2.38}$$

If one now decomposes ϕ_m into pieces which are transversal and longitudinal with respect to the operator $\mathring{\mathscr{D}}_m$, viz.

$$\psi_m = \chi_m + \mathring{\mathscr{D}}_m \chi \,, \quad \mathring{\mathscr{D}}^m \chi_m = 0 \,, \tag{2.39}$$

one infers from (2.38) that this decomposition is maintained by the spin-$\frac{1}{2}$ mass operator (note that the "de Sitter mass term" in (2.35) is essential here). Therefore, the eigenspace of this operator decomposes naturally into the space of spurious states with $\mathring{\mathscr{D}}^m \psi_m \neq 0$ and the space of physical states which obey

$$\mathring{\mathscr{D}}^m \psi_m = 0 \,. \tag{2.40}$$

Eq. (2.40) is a genuine seven-dimensional gauge condition. It should also be noted that the analogous decomposition

$$\psi_m = \chi'_m + \Gamma_m \chi' \,, \quad \Gamma^m \chi'_m = 0 \,, \tag{2.41}$$

is *not* invariant with respect to the spin-$\frac{1}{2}$ mass operator.

The eigenmodes of type (2.24) are now easy to identify: one simply expands the spinor χ in (2.39) into the complete set (2.33) of eigenmodes of the Dirac operator. The corresponding eigenvalues are then related to the eigenvalues λ of (2.35) by eq. (2.38). In this way, we find

$$\lambda = k + \tfrac{5}{2}\,, \quad \lambda = -k - \tfrac{7}{2} \quad (k \geq 1)\,. \tag{2.42}$$

Observe that the mode associated with the eigenvalue $\lambda = \frac{5}{2}$ is absent because, in this case, χ in (2.39) is the covariantly constant spinor which is annihilated by the operator $\overset{\circ}{\mathscr{D}}_m$; the wave function for ψ_m thus vanishes identically.

Having identified the modes of type (2.24), one next simplifies the ansatz (2.36) by rendering it orthogonal to these modes. This is equivalent to imposing the condition (2.40), and after a little calculation, (2.36) is replaced by

$$\begin{aligned}\psi_m^{\text{physical}} &= \alpha m_7^2 Y_m \eta + \beta m_7 Y^{p;n} \Gamma_{mnp} \eta + \gamma (m_7^2 Y^n \Gamma_{mn} \eta + m_7 Y_{m;n} \Gamma^n \eta) \\ &\quad + \delta (Y^{p;n}{}_{;m} \Gamma_{np} \eta - 5 m_7 Y_{n;m} \Gamma^n \eta - k(k+6) m_7 Y_{m;n} \Gamma^n \eta)\,. \end{aligned} \tag{2.43}$$

In this way, we have been able to eliminate two out of the six coefficients present in (2.36) and to reduce the determination of the physical eigenvalues to the computation of a 4 by 4 determinant. The relevant 4 by 4 matrix can be obtained by inserting the ansatz (2.43) into the left-hand side of (2.35); it is

$$\mathscr{M} = \begin{bmatrix} \frac{5}{2} - \lambda & 0 & k(k+6)+5 & -(k(k+6)^2+25) \\ 0 & -\frac{11}{2} - \lambda & 2 & 3 - k(k+6) \\ 1 & k(k+6)+5 & -\frac{7}{2} - \lambda & 0 \\ 0 & 1 & 0 & -\frac{7}{2} - \lambda \end{bmatrix}. \tag{2.44}$$

From (2.44), one calculates the eigenvalues which are given by

$$\lambda = k + \tfrac{7}{2} \quad (k \geq 1)\,, \quad -k - \tfrac{5}{2} \quad (k \geq 2)\,, \tag{2.45}$$

$$\lambda = k - \tfrac{5}{2}\,, \quad -k - \tfrac{17}{2} \quad (k \geq 1)\,. \tag{2.46}$$

The restriction to $k \geq 2$ in (2.45) requires some explanation. Substituting the putative eigenvalue $\lambda = -\frac{7}{2}$ (i.e. $k = 1$) into (2.44) and making use of the ansatz (2.43), we get the corresponding wave function

$$\begin{aligned}\psi_m &= 2(m_7^2 Y^n \Gamma_{nm} \eta + m_7 Y_{m;n} \Gamma^n \eta) \\ &\quad + Y^{p;n}{}_{;m} \Gamma_{np} \eta - 5 m_7 Y_{n;m} \Gamma^n \eta - 7 m_7 Y_{m;n} \Gamma^n \eta\,. \end{aligned} \tag{2.47}$$

The vector spherical harmonic Y_m for $k = 1$ may be explicitly represented by $\bar{\eta}^I \Gamma_m \eta^J$ in terms of covariantly constant spinors η^I, from which it follows that

$$Y_{m;n} = -Y_{n;m}\,, \quad Y^{p;n}{}_{;m} = 2 m_7^2 Y^{[p} \delta^{n]}_m\,. \tag{2.48}$$

Inserting (2.48) into (2.47), one sees that the wave function (2.47) vanishes identically and therefore the eigenvalue $\lambda = -\frac{7}{2}$ in (2.42) is, in fact, absent. Finally, it can be shown by explicit calculation, that the two towers (2.45), in addition to (2.40), satisfy the constraint $\Gamma^m \chi_m = 0$ whereas the other two towers (2.46) do not. It should, of course, be understood that the eigenmodes corresponding to (2.39), (2.42), (2.45) and (2.46) belong to irreducible representations of SO(8); we have suppressed their representation labels for simplicity. The symmetry assignments will be discussed in the following section.

We now briefly return to the gauge condition (2.12). It has already been pointed out that it connects four- and seven-dimensional quantities but its implications in the four-dimensional context must still be elucidated. For the modes (2.42), one obviously has $\mathring{\mathscr{D}}^m \psi_m \neq 0$ and therefore the gauge condition (2.27) determines a constraint on $\mathring{\mathscr{D}}^\mu \psi_\mu$ for the four-dimensional spin-$\frac{3}{2}$ fields. More precisely, it follows from (2.42) that there is an exact correspondence between the massive gravitinos and these modes which furthermore belong to the *same* representations of SO(8). Thus, through the constraint on $\mathring{\mathscr{D}}^\mu \psi_\mu$, the modes (2.42) provide the required helicity-$\frac{1}{2}$ states to make the spin-$\frac{3}{2}$ fields massive. (This correspondence is also useful in the group theoretical treatment.) On the other hand, the massless gravitino has no associated spurious mode, as was mentioned after eq. (2.42), and the gauge condition (2.11) reads

$$\mathring{\mathscr{D}}^\mu \psi_\mu = 0 \tag{2.49}$$

in this case. Eq. (2.49) eliminates the helicity-$\frac{1}{2}$ degree of freedom and expresses the masslessness of the lowest-lying gravitino as well as the existence of eight supersymmetries of the ground state.

The masses of the spin-$\frac{1}{2}$ and spin-$\frac{3}{2}$ particles are defined by the eigenvalues $m_{1/2}$ and $m_{3/2}$ of the four-dimensional differential operators appearing in the left-hand side of eqs. (2.28) and (2.29). They are given, in units of m_7, by eqs. (2.33), (2.45) and (2.46) up to an additive constant. We thus have

$$\begin{aligned}
m_{3/2}^{(1)} &= k+2\,, & k &\geq 0\,,\\
m_{3/2}^{(2)} &= -k-4\,, & k &\geq 1\,,\\
m_{1/2}^{(1)} &= k-1\,, & k &\geq 1\,,\\
m_{1/2}^{(2)} &= -k-7\,, & k &\geq 1\,,\\
m_{1/2}^{(3)} &= k+5\,, & k &\geq 1\,,\\
m_{1/2}^{(4)} &= -k-1\,, & k &\geq 2\,.
\end{aligned} \tag{2.50}$$

The superscripts label the towers. The members of the "massless" $N=8$ supermultiplet are at the bottom of the towers $m_{3/2}^{(1)}$ and $m_{1/2}^{(1)}$. Note that for convenience the "massless" gravitino has been given the value +2. We do not list the SO(8) content of the modes here as this will be discussed in the next section.

3. Osp(8, 4) classification

Up to this point, the symmetry assignments of the various modes have not been discussed in any detail. The mass spectrum of $N=8$ supergravity on S^7 in the bosonic and fermionic case has been determined by solving the appropriate eigenvalue equations, and no explicit reference to the SO(8) and supersymmetry content of the modes was necessary. The SO(8) assignments can be deduced from those of the spherical (scalar, vector and tensor) harmonics on S^7 which are known [18], but this is not sufficient to group the various states into supermultiplets. For a complete classification, one has to make use of the full invariance of the S^7 ground state. This group contains not only the 28 rotations of SO(8) corresponding to the 28 Killing spinors on S^7 but also eight spinorial translations which correspond to the eight Killing spinors on S^7. Together, the generators associated with these bosonic and fermionic transformations constitute the graded Lie algebra Osp(8, 4), and a rigorous proof of the Osp(8, 4) invariance of the S^7 ground state has been given in ref. [20]. The excitations corresponding to the fluctuations about the ground state should therefore form irreducible representations of Osp(8, 4). From the general Kaluza–Klein theory [8] and the absence of higher spin fields in eleven-dimensional supergravity, it follows that the relevant representations are those with maximum spin 2. The latter have been classified in ref. [21]; and we will restrict our attention to these representations here.

The masses of the excited states are proportional to the inverse radius $|m_7|$ of the seven sphere. Thus, in the limit $m_7 \to 0$ where the space becomes flat, all masses tend to zero. In this limit, the relevant superalgebra is the Poincaré superalgebra, and we conclude that in this contraction limit, the massive representations of Osp(8, 4) become *massless* representations of $N=8$ Poincaré supersymmetry. This has the very important consequence that all massive representations of Osp(8, 4) with maximum spin 2 must be obtainable from massless representations of $N=8$ supersymmetry with the same spin limit. There is only one such multiplet with maximum spin 2, namely the massless $N=8$ multiplet already mentioned in the introduction. It contains one graviton [$\underline{1}$ of SO(8)], eight gravitinos ($=8_s$), 28 spin-1 fields ($=28$), 56 spin-$\frac{1}{2}$ fields ($=56_s$), 35 scalars ($=35_v$) and 35 pseudoscalars ($=35_c$) (for the group theoretic conventions, see ref. [22]). Hence, one should be able to derive all massive Osp(8, 4) multiplets from products of the form

$$R \otimes \{1, 8_s, 28, 56_s, 35_v, 35_c\}\,, \tag{3.1}$$

where R is an as yet unspecified representation of SO(8).

To facilitate the discussion, we next introduce Dynkin labels to classify the representations of SO(8) [22]. Each irreducible representation of SO(8) can be uniquely labelled by a set $(a_1a_2a_3a_4)$ of four non-negative integers a_1, a_2, a_3, a_4. Since the massless graviton which belongs to the massless $N=8$ multiplet is an SO(8) singlet, the charged massive gravitons will carry the same label as the relevant irreducible representation. One now realizes that the representation R which occurs in (3.1) is no longer arbitrary, since we know from the explicit calculations [8] that the massive gravitons are in one-to-one correspondence with the eigenfunctions of the laplacian on S^7, i.e. the spherical harmonics on S^7. These are characterized by the Dynkin labels $(n000)$, $n \in N$, which correspond to the symmetric and traceless

SO(8) tensors with n indices. To obtain the full Osp(8, 4) multiplet, we replace R in (3.1) by $n(n000)$, perform the multiplication and identify the irreducible components in this product. The Dynkin labels of the massless representation are given by

$$8_s = (0001)\,, \quad 28 = (0100)\,,$$
$$56_s = (1010)\,, \quad 35_v = (2000)\,, \quad 35_c = (0020)\,. \tag{3.2}$$

The result of this multiplication, which is given in ref. [21], is, however, not yet the final answer. One still has to add lower helicity states to the spin-2, spin-$\frac{3}{2}$ and spin-1 fields to make them massive. The lower helicity states which are absorbed must belong to the same representation as the gauge field into which they are absorbed. The final result which is obtained after absorbing these states reads*:

$$\begin{aligned}
&\text{spin-2:} && (n000)\,,\\
&\text{spin-}\tfrac{3}{2}\text{:} && (n001)\oplus(n-1010)\,,\\
&\text{spin-}1^+\text{:} && (n-1011)\,,\\
&\text{spin-}1^-\text{:} && (n100)\oplus(n-2100)\,,\\
&\text{spin-}\tfrac{1}{2}\text{:} && (n+1010)\oplus(n-1110)\oplus(n-2101)\oplus(n-2001)\,,\\
&\text{spin-}0^+\text{:} && (n+2000)\oplus(n-2200)\oplus(n-2000)\,,\\
&\text{spin-}0^-\text{:} && (n020)\oplus(n-2002)\,,
\end{aligned} \tag{3.3}$$

where, whenever an integer is negative, the associated representation does not exist; for example, the second spin-$\frac{3}{2}$ tower starts only at $n=1$. For each n, (3.2) is an irreducible representation of Osp(8, 4), and the integer n therefore labels the "floors" of the massive tower.

To relate the group theoretical result (3.3) to the solutions of the eigenvalue equations of the preceding section, one must properly adjust the relation between the index n which labels Osp(8, 4) multiplets and the index k which labels the spherical harmonics. For the spin-2 and spin-$\frac{3}{2}$ states, the identification is straightforward; for example, the eigenmodes of the Dirac operator found in (2.32) and (2.33) exactly correspond to the two representations in (3.3), and the absence of the "ground floor" for the second tower was also obtained there. The spurious modes (2.42) which are eaten by the gravitinos also appear in the product (3.1) but have already been absorbed in (3.3). For the spin-$\frac{1}{2}$ states, one makes use of the decomposition

$$(l100)\otimes(0001) = (l101)\oplus(l+1010)\oplus(l-1110)\oplus(l001)\,, \tag{3.4}$$

which yields the physical modes because the vector spherical harmonics Y_m with index $k \geqslant 1$ belong to the representations $(l100)$ with $l = k-1 \geqslant 0$. The Γ traceless eigenmodes (2.29) then correspond to the $(l-2101)$ and $(l-1110)$ representations in (3.4). That these assignments are indeed correct follows from the gauge condition (2.25): the term $\gamma^\mu\psi_\mu$ can only belong to $(l001)$ or $(l010)$, and any representation for ψ_m which is different must satisfy the constraint $\Gamma^m\psi_m=0$. Similar considerations apply to the spin-0 sector.

---

* For the special case $n=1$, this result was first obtained in ref. [19].

Besides spin and SO(8) content, the complete characterization of the Osp(8, 4) states requires the knowledge of the lowest eigenvalues E_0 of the "energy operator" M_{04} of the SO(2, 3) subalgebra of Osp(8, 4). The energy labels are most easily found using the fact that

$$\mathrm{Osp}(8,4)\supset \mathrm{Osp}(1,4)\times \mathrm{SO}(7) \tag{3.5}$$

and the known energy labels of Osp(1, 4) representations [23]. To illustrate this procedure, we first analyze the $n=0$ multiplet (3.2). Under the SO(7) subgroup of SO(8), these representations decompose as $1\to 1$; $8_s\to 1+7$, $28\to 7+21$; $56\to 21+35$; $35_v\to 35$ and $35_c\to 35$. It is evident that the members of an Osp(1, 4) multiplet must belong to the same representation of SO(7); on the other hand, states that emerge from the same SO(8) representation must carry the same energy label. The 35 corresponds to a Wess–Zumino multiplet in $(\mathrm{AdS})_4$, and since it is massless, it is uniquely characterized by [23]:

$$D(1,0)\oplus D(\tfrac{3}{2},\tfrac{1}{2})\oplus D(2,0)\,, \tag{3.6}$$

where the first number is the energy E_0 and the second the spin s of the lowest state in the SO(2, 3) multiplet. Massless higher spin representations are characterized by $(s=\tfrac{1}{2},1,\ldots)$

$$D(s+1,s)\oplus D(s+\tfrac{3}{2},s+\tfrac{1}{2})\,. \tag{3.7}$$

Below, we also need the general massive higher spin representations of Osp(1, 4), which are given by

$$D(E_0,s)\oplus D(E_0+\tfrac{1}{2},s+\tfrac{1}{2})\oplus D(E_0+\tfrac{1}{2},s-\tfrac{1}{2})\oplus D(E_0+1,s)\,, \tag{3.8}$$

where $E_0>s+1$, $s=\tfrac{1}{2},1,\ldots$. Putting everything together, we obtain table 1.

TABLE 1

The $n=0$ Osp(8, 4) multiplet

Spin	2	$\frac{3}{2}$	1^-	$\frac{1}{2}$	0^-	0^+
SO(8)	1	8_s	28	56_s	35_c	35_c
SO(7)	1	1				
decomposition		7	7			
			21	21		
				35	35	35
E_0	3	$\frac{5}{2}$	2	$\frac{3}{2}$	2	1

TABLE 2

The $n=1$ Osp(8, 4) multiplet

Spin	2	$\frac{3}{2}$	$\frac{3}{2}$	1^+	1^-	$\frac{1}{2}$	$\frac{1}{2}$	0^-	0^+
SO(8)	8_v	8_c	56_c	56_v^+	160_v^-	160_c	224_{vc}	224_{cv}	112_v
SO(7)	8	8	8	8					
decomposition			48	48	48	48			
					112	112	112	112	
							112	112	112
E_0	$\frac{7}{2}$	4	3	$\frac{7}{2}$	$\frac{5}{2}$	3	2	$\frac{5}{2}$	$\frac{3}{2}$

A similar analysis for the $n=1$ Osp(8, 4) multiplet leads to the results shown in table 2.

This construction is easily generalized to the higher excited multiplets on S^7. The result is given by:

$$
\begin{aligned}
s=2:&\quad E_0(n000)=3+\tfrac{1}{2}n\,,\\
s=\tfrac{3}{2}:&\quad E_0(n001)=\tfrac{5}{2}+\tfrac{1}{2}n\,,\\
&\quad E_0(n-1010)=\tfrac{7}{2}+\tfrac{1}{2}n\,,\\
s=1:&\quad E_0(n100)=2+\tfrac{1}{2}n\,,\\
&\quad E_0(n-1011)=3+\tfrac{1}{2}n\,,\\
&\quad E_0(n-2100)=4+\tfrac{1}{2}n\,,\\
s=\tfrac{1}{2}:&\quad E_0(n+1010)=\tfrac{3}{2}+\tfrac{1}{2}n\,,\\
&\quad E_0(n-1110)=\tfrac{5}{2}+\tfrac{1}{2}n\,,\\
&\quad E_0(n-2101)=\tfrac{7}{2}+\tfrac{1}{2}n\,,\\
&\quad E_0(n-2001)=\tfrac{9}{2}+\tfrac{1}{2}n\,,\\
s=0:&\quad E_0(n+2000)=1+\tfrac{1}{2}n\,,\\
&\quad E_0(n020)=2+\tfrac{1}{2}n\,,\\
&\quad E_0(n-2200)=3+\tfrac{1}{2}n\,,\\
&\quad E_0(n-2002)=4+\tfrac{1}{2}n\,,\\
&\quad E_0(n-2000)=5+\tfrac{1}{2}n\,,
\end{aligned}
\tag{3.9}
$$

from which one reads off the *universal* mass energy relation

$$E_0=\tfrac{3}{2}+\tfrac{1}{2}\sqrt{m^2+1}\quad\text{for bosons}\,, \tag{3.10}$$

$$E_0=\tfrac{3}{2}+\tfrac{1}{2}|m|\quad\text{for fermions}\,. \tag{3.11}$$

Eq. (3.10) is valid for all bosonic states except the 35_v massless scalars for which

TABLE 3

The spectrum of supergravity on the seven-sphere

Spin	SO(8) content		$(\text{Mass})^2$ in units of m_7^2
2^+	$(n000)^*$	$n \geqslant 0$	$(n+3)^2-1$
$\frac{3}{2}^{(1)}$	$(n0001)^*$	$n \geqslant 0$	$(n+2)^2$
$\frac{3}{2}^{(2)}$	$(n-1010)$	$n \geqslant 1$	$(n+4)^2$
$1^{-(1)}$	$(n100)^*$	$n \geqslant 0$	$(n+1)^2-1$
1^+	$(n-1011)$	$n \geqslant 1$	$(n+3)^2-1$
$1^{-(2)}$	$(n-2100)$	$n \geqslant 2$	$(n+5)^2-1$
$\frac{1}{2}^{(1)}$	$(n+1010)^*$	$n \geqslant 0$	n^2
$\frac{1}{2}^{(2)}$	$(n-1110)$	$n \geqslant 1$	$(n+2)^2$
$\frac{1}{2}^{(3)}$	$(n-2101)$	$n \geqslant 2$	$(n+4)^2$
$\frac{1}{2}^{(4)}$	$(n-2001)$	$n \geqslant 2$	$(n+6)^2$
$0^{+(1)}$	$(n+2000)^*$	$n \geqslant 0$	$(n-1)^2-1$
$0^{-(1)}$	$(n020)^*$	$n \geqslant 0$	$(n+1)^2-1$
$0^{+(2)}$	$(n-2200)$	$n \geqslant 2$	$(n+3)^2-1$
$0^{-(2)}$	$(n-2002)$	$n \geqslant 2$	$(n+5)^2-1$
$0^{+(3)}$	$(n-2000)$	$n \geqslant 2$	$(n+7)^2-1$

The states marked by an asterisk contain the zero-mass supermultiplet.

we have

$$E_0 = 1 = \tfrac{3}{2} - \tfrac{1}{2}\sqrt{m^2+1}\,. \tag{3.12}$$

Collecting all our results we get table 3.

The universality of eqs. (3.10) and (3.11) implies that E_0 has a dynamical significance. In fact, we know from ref. [24] that the relation (3.10) for spin-0 fields characterizes modes which die fast enough at infinity to ensure energy conservation in AdS. The reality of E_0 is guaranteed by the fact that $m^2 \geqslant -1$, the stability limit being reached in the $0^{+(1)}$ tower for $n=1$ ($k=3$) with a multiplet of 112 scalars. For $n=2$ ($k=4$) the same tower again contains conformal massless modes (294 scalars) with vanishing energy flow at spatial infinity. However, they must satisfy different boundary conditions, characterized by the + sign in eq. (3.10), than the 35_v in order not to break supersymmetry. In this way they fit indeed as massless members in the "massive" supermultiplet $n=2$. Note that, in contrast to Poincaré supersymmetry, states belonging to the same supermultiplet characterized by n may have different masses because of the non-commutativity of the energy operator M_{04} with supersymmetry generators. Thus we see that for scalar modes, the supersymmetric spectrum is consistent with a Hilbert space of functions with a boundary condition preventing energy flow in and out of AdS and hence admitting well-defined Cauchy data in this otherwise unviable space. We infer that such a property holds for all the modes because of the universality of eqs. (3.10) and (3.11), a conjecture that could be checked explicitly following the method of ref. [24]. It follows from the reality of

E_0 that the seven-sphere is stable against small fluctuations belonging to this Hilbert space, a fact which also follows from supersymmetry [25].

Finally, we remark that the group theoretical calculation based on the analysis of Osp(8, 4) multiplets which leads to (3.9) is essentially *quantum mechanical.* In contrast, the method used in sect. 2, which is based on harmonic expansions on S^7, yields the *classical mass spectrum* only. The agreement between the final results motivates the conjecture that the results of table 3 are, in fact, valid to all orders in perturbation theory, if the quantization procedure respects $N = 8$ supersymmetry.

References

[1] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409
[2] J. Scherk and J. Schwarz, Phys. Lett. B57 (1975) 463;
E. Cremmer and J. Scherk, Nucl Phys. B103 (1976) 393; B108 (1976) 409
[3] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233
[4] B. Biran, B. de Wit, F. Englert and H. Nicolai, Phys. Lett. 124B (1983) 45
[5] E. Cremmer and B. Julia, Phys. Lett. 80B (1978) 48; Nucl. Phys. B159 (1979) 141
[6] M.J. Duff and C.N. Pope, in Supersymmetry and supergravity '82, eds. S. Ferrara, J.G. Taylor and P. van Nieuwenhuizen (World Scientific, 1983)
[7] B. de Wit and H. Nicolai, Phys. Lett. 108B (1981) 285; Nucl. Phys. B208 (1982) 232
[8] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Kl (1921) 966;
O. Klein, Z. Phys. 37 (1926) 895;
B. De Witt, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965);
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautmann, Rep. Math. Phys. 1 (1970) 29;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
C. Orzalesi, Fortschr. Phys. 29 (1981) 413;
E. Witten, Nucl. Phys. B186 (1981) 412;
A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316
[9] M.J. Duff and D.J. Toms, in Unification of the fundamental interactions II, eds. S. Ferrara and J. Ellis (Plenum, 1982)
[10] F. Englert and H. Nicolai, CERN Preprint TH. 3711 (1983), to appear in Proc. XII Int. Colloq. on Group theoretical methods in physics, Trieste (1983)
[11] B. Biran, A. Casher, F. Englert, M. Rooman and P. Spindel, Phys. Lett. 134B (1984) 179
[12] E. Sezgin, preprint IC/83/220 (1983)
[13] R. D'Auria and P. Fré, Torino preprint IFTT430 (1983)
[14] M.A. Awada, M.J. Duff and C.N. Pope, Phys. Rev. Lett. 50 (1983) 294
[15] F.A. Bais, H. Nicolai and P. van Nieuwenhuizen, Nucl. Phys. B228 (1983) 333
[16] B.E.W. Nilsson and C.N. Pope, Imperial College preprint ICTP/82–83/27 (1983)
[17] K. Yamagishi, Mass spectrums of vector particles in $D = 11$ supergravity compactified on squashed seven sphere, Osaka preprint (1983)
[18] S. Gallot and D. Meyer, J. Math. Pures et Appl. 54 (1975) 259;
G.W. Gibbons and M.J. Perry, Nucl. Phys. B146 (1978) 90;
Bateman manuscript, vol. II (McGraw Hill, New York, 1953) p. 232
[19] M.J. Duff, B.E.W. Nilsson and C.N. Pope, Phys. Rev. Lett. 50 (1983) 2043
[20] R. d'Auria and P. Fré, Phys. Lett. 121B (1983) 141
[22] D.Z. Freedman and H. Nicolai, Nucl. Phys. B237 (1984) 342
[22] R. Slansky, Phys. Reports C79 (1981) 1
[23] W. Heidenreich, Phys. Lett. 110B (1982) 461
[24] P. Breitenlohner and D.Z. Freedman, Phys. Lett. 115B (1982) 97; Ann. of Phys. 144 (1982) 249
[25] G.W. Gibbons, C.M. Hull and N.P. Warner, Nucl. Phys. B128 (1983) 173

$N = 8$ Supergravity Breaks Down to $N = 1$

M. A. Awada, M. J. Duff,[(a)] and C. N. Pope
Blackett Laboratory, Imperial College, London SW7 2BZ, England

(Received 9 December 1982)

The field equations of $N = 1$ supergravity in $d = 11$ dimensions admit a spontaneous compactification on the seven-sphere to an $N = 8$ theory in $d = 4$ with local SO(8) invariance. A spontaneously broken version of this theory is provided by another solution of the field equations for which the metric on S^7 is distorted. The isometry group of this new solution is $SO(5) \otimes SU(2)$, the relevant holonomy group is G_2, and the $N = 8$ supersymmetry is broken down to $N = 1$ at the Planck scale. The implications for grand unified theories are briefly discussed.

PACS numbers: 04.60.+n, 11.30.Pb, 12.25.+e

Since its discovery in 1976, progress in supergravity[1] has evolved along rather diverse lines. There are those who, awed by the majesty of extended supersymmetry, have looked to the $N = 8$ theory as a possible unification of all interactions, while the more phenomenologically minded have concentrated on $N = 1$ supersymmetry as a means of solving the gauge hierarchy problem and accomodating chiral fermion representations in grand unified theories. Consequently, the hope has sometimes been expressed that these two approaches could be linked if $N = 8$ supersymmetry were to break down to $N = 1$ at the Planck scale. In this paper, we demonstrate that this hope can indeed be realized.

Our starting point is the observation[2,3] that the field equations of $N = 1$ supergravity in $d = 11$ dimensions[4] admit a candidate ground-state solution corresponding to the product of four-dimensional anti–de Sitter space (AdS) and the seven-sphere with its standard metric, i.e., the coset space SO(8)/SO(7). This seven-sphere admits 28 Killing vectors and eight Killing spinors (i.e., eight spinors which are covariantly constant with respect to the de Sitter covariant derivative appearing in the transformation law for the gravitino). It gives rise, à la Kaluza-Klein, to an effective four-dimensional theory with local SO(8) invariance and $N = 8$ supersymmetry. The SO(8) Yang-Mills coupling constant e is given by $e^2 \sim m^2 M_P^{-2}$, where m^{-1} is the radius of S^7 and M_P is the Planck mass. Thus m is of order M_P if e is of order 1. The massless sector describes one spin 2, eight spin $\frac{3}{2}$, 28 spin 1, 56 spin $\frac{1}{2}$, and 70 spin 0 (35 scalars plus 35 pseudoscalars) and may probably be identified with the gauged $N = 8$ theory of de Wit and Nicolai,[5] where, in addition to the obvious local SO(8) with 28 elementary gauge bosons, one finds a hidden SU(8) with 63 composite gauge bosons. This SU(8) has formed the basis for possible grand unification schemes.[6]

By extending some earlier work on "squashing" three-spheres and symmetry breaking,[7] it has recently been suggested[2,3] that other solutions of the $d = 11$ field equations which are topologically still S^7 but which deviate from the maximally symmetric S^7 geometry could give rise to a spontaneously broken version of this gauged $N = 8$ theory. The idea is that distortion of the seven-sphere corresponds to nonvanishing vacuum expectation values for the scalar fields and hence to a Higgs and super-Higgs effect in $d = 4$. In this paper we exhibit such a solution corresponding to the fact that S^7 admits not one but two Einstein metrics. In addition to the "round" S^7 discussed above there is a "squashed" S^7 with isometry group $Sp(2) \otimes Sp(1) \cong SO(5) \otimes SU(2)$ and for which the de Sitter connection has holonomy group G_2. Remarkably, we find that it admits but one Killing spinor and hence yields an effective four-dimensional theory with one unbroken supersymmetry. Thus $N = 8$ supersymmetry is broken down to $N=1$ at the Planck scale.

We adopt the following conventions. The $d=11$ space has signature $(-+++\ldots)$, and $d=11$ indices $M, N, \ldots$ will be decomposed into $d=4$ indices $\mu, \nu, \ldots$ and $d=7$ indices $m, n, \ldots$. The $d=11$ Dirac matrices satisfy

$$\{\Gamma_A, \Gamma_B\}=2\eta_{AB} \tag{1}$$

and may be written

$$\Gamma_A=(\gamma_\alpha\otimes\underline{1},\ \gamma_5\otimes\Gamma_a), \tag{2}$$

where γ_α and Γ_a are the Dirac matrices in $d=4$ and $d=7$, respectively. Spontaneous compactification works as follows: We set the fermion fields to zero and examine the boson field equation of the $d=11$ theory[4]:

$$R_{MN}-\tfrac{1}{2}g_{MN}R=\tfrac{1}{3}[F_{MPQR}F_N{}^{PQR}-\tfrac{1}{8}g_{MN}F^2], \tag{3}$$

$$\nabla_M F^{MNPQ} = -\tfrac{1}{576}\epsilon^{M_1\ldots M_8NPQ}F_{M_1\ldots M_4}F_{M_5\ldots M_8}. \tag{4}$$

Equation (4) is solved by the Freund-Rubin[8] choice for which all components F_{MNPQ} vanish except

$$F_{\mu\nu\rho\sigma}=3m\epsilon_{\mu\nu\rho\sigma}, \tag{5}$$

where m is a constant. Substituting into Eq. (3) gives

$$R_{\mu\nu}=-12m^2g_{\mu\nu} \tag{6}$$

and

$$R_{mn}=6m^2g_{mn}. \tag{7}$$

Thus the eleven-dimensional space becomes a product of a four-dimensional Einstein space with negative cosmological constant and a seven-dimensional Einstein space with positive cosmological constant.

There are still infinitely many solutions of (6) and (7) but the ground state should presumably be distinguished by its symmetries. With this in mind, we proceed as in Refs. 2 and 3 to restrict the solution further by requiring that the vacuum be supersymmetric, i.e., by requiring the existence of covariantly constant spinors ϵ for which

$$\delta\psi_M=\overline{D}_M\epsilon=0, \tag{8}$$

where

$$\overline{D}_M=D_M-\frac{i}{144}(\Gamma^{NPQR}{}_M+8\Gamma^{PQR}\delta_M{}^N)F_{NPQR}\epsilon. \tag{9}$$

Substituting Eq. (5) into Eq. (9) yields

$$\overline{D}_\mu=D_\mu+m\gamma_\mu\gamma_5, \tag{10}$$

$$\overline{D}_m=D_m-\tfrac{1}{2}m\Gamma_m. \tag{11}$$

When $m\neq 0$, the requirement of $N=8$ supersymmetry singles out the unique choice of AdS$\otimes S^7$ with the standard SO(8)-invariant metric on S^7. This is because there will be one unbroken supersymmetry (i.e., one massless gravitino) for each Killing spinor in $d=7$, i.e., each spinor satisfying

$$\overline{D}_m\eta=0. \tag{12}$$

The integrability condition is

$$[\overline{D}_m, \overline{D}_n]\eta=-\tfrac{1}{4}R_{mn}{}^{ab}\Gamma_{ab}\eta+\tfrac{1}{2}m^2\Gamma_{mn}\eta. \tag{13}$$

Since spinors in $d=7$ have eight components, there will be eight covariantly constant spinors if and only if

$$R_{mnpq}=m^2(g_{mp}g_{nq}-g_{mq}g_{np}), \tag{14}$$

and we obtain the maximally symmetric special case of Eq. (7) corresponding to the seven-sphere with its standard metric of constant curvature. Similarly the vanishing of $[\overline{D}_\mu, \overline{D}_\nu]$ on spinors in $d=4$ implies that space-time is AdS = SO(3, 2)/SO(3, 1). As discussed in Refs. 2 and 3, this vacuum solution then yields the effective $d=4$ theory with local SO(8) invariance and $N=8$ supersymmetry.

One can now contemplate other solutions of Eq. (7), i.e., other seven-dimensional Einstein metrics which might admit fewer than eight Killing spinors. Of particular interest would be one with the same S^7 topology since this would correspond to a spontaneously broken version of the previous theory. Remarkably S^7 does indeed admit another Einstein metric for which the sphere is squashed in a special homogeneous manner. It may be described as the distance sphere in $P_2(H)$ the quaternionic projective plane, i.e., as the level surface formed by all geodesics of length r emanating from a point in $P_2(H)$.[9] When r is very small, the distance sphere approaches the standard "round" S^7 [which is an S^3 bundle over S^4; in fact this corresponds to the $K=1$ SU(2) Yang-Mills instanton in four-dimensional Euclidean space] but as r increases the length of the S^3 fibers shrinks relative to the size of the S^4 base, and the eigenvalues α_n $(0\leq n\leq 6)$ of the Ricci tensor split into two sets $\alpha_0=\alpha_1=\alpha_2=\alpha_3=\alpha$, $\alpha_4=\alpha_5=\alpha_6=\beta$, $\alpha\neq\beta$. However, for a certain value of r the two sets become equal again, and the metric becomes an Einstein metric. The symmetry of this distorted sphere is the group which leaves a point in $P_2(H)$ fixed, namely $\mathrm{Sp}(2) \otimes \mathrm{Sp}(1)$ [where $\mathrm{Sp}(n)$ is the group of $n \times n$ quaternionic matrices], which is isomorphic to $\mathrm{SO}(5) \otimes \mathrm{SU}(2)$.

Explicitly, the Fubini-Study Einstein metric on $P_2(H)$, written in terms of quaternionic coordinates (q_1, q_2), is

$$ds^2 = (1 + \sum_k \bar{q}_k q_k)^{-1} \sum_i d\bar{q}_i dq_i - (1 + \sum_k \bar{q}_k q_k)^{-2} \sum_{i,j} \bar{q}_i dq_i d\bar{q}_j q_j . \tag{15}$$

Defining

$$q_1 = \tan\chi \cos(\tfrac{1}{2}\mu) U, \quad q_2 = \tan\chi \sin(\tfrac{1}{2}\mu) V, \tag{16}$$

where U and V are quaternions of unit modulus, it is straightforward to show that the 1-forms σ_i, Σ_i defined by

$$U^{-1} dU = i\sigma_1 + j\sigma_2 + k\sigma_3, \quad V^{-1} dV = i\Sigma_1 + j\Sigma_2 + k\Sigma_3, \tag{17}$$

satisfy the algebra $d\sigma_1 = -\sigma_2 \wedge \sigma_3$, etc., of SU(2). ($i, j, k$ are the imaginary quaternions.) Finally, defining

$$\nu_i = \sigma_i + \Sigma_i, \quad \omega_i = \sigma_i - \Sigma_i, \tag{18}$$

the metric on $P_2(H)$ becomes

$$ds^2 = d\chi^2 + \tfrac{1}{4}\sin^2\chi [d\mu^2 + \tfrac{1}{4}\sin^2\mu\, \omega_i{}^2 + \tfrac{1}{4}\cos^2\chi (\nu_i + \cos\mu\, \omega_i)^2]. \tag{19}$$

The distance sphere of radius $r (= \tan\chi)$, centered on $q_1 = q_2 = 0$, inherits the metric obtained from (19) by setting $\chi = \text{const}$. Thus up to an overall scaling constant, the squashed seven-sphere has metric

$$ds^2 = d\mu^2 + \tfrac{1}{4}\sin^2\mu\, \omega_i{}^2 + \tfrac{1}{4}\lambda^2 (\nu_i + \cos\mu\, \omega_i)^2, \tag{20}$$

where λ is a constant parameter describing the degree of distortion. The round, SO(8)-invariant, sphere corresponds to $\lambda^2 = 1$.

Introducing the orthonormal basis $e^0 = d\mu$, $e^1 = \tfrac{1}{2}\sin\mu\, \omega_1, \ldots, e^4 = \tfrac{1}{2}\lambda(\nu_1 + \cos\mu\, \omega_1)$, etc., a straightforward calculation shows that the curvature form $\theta_{mn} = \tfrac{1}{2} R_{mnpq} e^p \wedge e^q$ is given by

$$\begin{aligned}
\theta_{01} &= (1 - \tfrac{3}{4}\lambda^2) e^0 \wedge e^1 + \tfrac{1}{2}(1 - \lambda^2) e^5 \wedge e^6, \\
\theta_{04} &= \tfrac{1}{4}\lambda^2 e^0 \wedge e^4 + \tfrac{1}{4}(1 - \lambda^2)[e^3 \wedge e^5 - e^2 \wedge e^6], \\
\theta_{12} &= (1 - \tfrac{3}{4}\lambda^2) e^1 \wedge e^2 + \tfrac{1}{2}(1 - \lambda^2) e^4 \wedge e^5, \\
\theta_{14} &= \tfrac{1}{4}\lambda^2 e^1 \wedge e^4 + \tfrac{1}{4}(1 - \lambda^2)[e^3 \wedge e^6 + e^2 \wedge e^5], \\
\theta_{15} &= \tfrac{1}{4}\lambda^2 e^1 \wedge e^5 - \tfrac{1}{4}(1 - \lambda^2)[e^0 \wedge e^6 + e^2 \wedge e^4], \\
\theta_{16} &= \tfrac{1}{4}\lambda^2 e^1 \wedge e^6 + \tfrac{1}{4}(1 - \lambda^2)[e^0 \wedge e^5 - e^3 \wedge e^4], \\
\theta_{45} &= \tfrac{1}{4}\lambda^{-2} e^4 \wedge e^5 + \tfrac{1}{2}(1 - \lambda^2)[e^0 \wedge e^3 + e^1 \wedge e^2],
\end{aligned} \tag{21}$$

where the remaining 14 components are obtained by performing simultaneous cyclic permutations of the triplets (1, 2, 3) and (4, 5, 6). The Ricci tensor $R_{mn} = \mathrm{diag}(\alpha, \alpha, \alpha, \alpha, \beta, \beta, \beta)$ with

$$\alpha = 3 - \tfrac{3}{2}\lambda^2, \quad \beta = \lambda^2 + 1/2\lambda^2 . \tag{22}$$

The Einstein condition, $\alpha = \beta$, gives two solutions for λ^2. $\lambda^2 = 1$ gives the round sphere discussed in Refs. 2 and 3, while $\lambda^2 = \tfrac{1}{5}$ gives the squashed sphere of this paper.

Substituting Eq. (21) with $\lambda^2 = \tfrac{1}{5}$ into Eq. (13) we find

$$[\bar{D}_m, \bar{D}_n] \eta = -\tfrac{1}{4} C_{mn} \eta, \tag{23}$$

where

$$C_{mn} = C_{mn}{}^{ab} \Gamma_{ab} \tag{24}$$

and $C_{mn}{}^{ab}$ is the Weyl tensor. Letting m run over $(0, i, \hat{i})$ where $i = 1, 2, 3$, and $\hat{i} = 4, 5, 6, = \hat{1}, \hat{2}, \hat{3}$, we find the 14 linearly independent components

$$C_{0i} = \tfrac{4}{5}[\Gamma_{0i} + \tfrac{1}{2}\epsilon_{ijk} \Gamma_{\hat{j}\hat{k}}], \tag{25}$$

$$C_{ij} = \tfrac{4}{5}[\Gamma_{ij} + \Gamma_{\hat{i}\hat{j}}], \tag{26}$$

$$C_{i\hat{j}} = \tfrac{4}{5}[-\Gamma_{i\hat{j}} - \tfrac{1}{2}\Gamma_{j\hat{i}} + \tfrac{1}{2}\delta_{ij}\Gamma_{k\hat{k}} - \tfrac{1}{2}\epsilon_{ijk}\Gamma_{0\hat{k}}], \tag{27}$$

where $C_{i\hat{j}}$ is trace free. The subgroup of SO(7) generated by these 14 linear combinations of Γ_{ab} corresponds to the holonomy[10] group of the connection of Eq. (11). Using the standard classification of Lie groups,[11] we find the exceptional group G_2.

To solve $C_{0i}\eta = 0 = C_{ij}\eta$ we resorted to an explicit representation,

$$\Gamma_0 = \gamma_0 \otimes \underline{1}, \quad \Gamma_i = \gamma_i \otimes \underline{1}, \quad \Gamma_{\hat{i}} = i\gamma_5 \otimes \tau_i, \qquad (28)$$

and found that there is one solution. It automatically satisfies $C_{i\hat{j}}\eta = 0$. The existence of just one Killing spinor is related to the fact that G_2 is the stability subgroup of a seven-dimensional spinor. Thus spontaneous compactification on the squashed S^7 gives an effective $d = 4$ theory with one unbroken supersymmetry. This means in particular that seven of the eight gravitinos will acquire masses of order M_P.

Carrying out the complete calculation of the four-dimensional Lagrangian will not be an easy task, however, and it remains to be seen what influence the isometry group of SO(5) ⊗ SU(2) and the holonomy group of G_2 will have on the classification of particle states. We simply note that both contain SU(2) ⊗ SU(2) as a subgroup. There is a subtle interplay between the obvious SO(8) symmetry and hidden SU(8) symmetry in the unbroken phase and we expect an analogous interplay in the spontaneously broken phase. It will be interesting to see what effect this might have on the SU(8) unification schemes. As discussed by Witten,[12] in addition to the gauge hierarchy problem and chiral fermion problem in grand unified theories, there is another reason why one unbroken supersymmetry at the energy of compactification is an attractive feature of Kaluza-Klein theories: It can provide a mechanism for stabilizing the Kaluza-Klein vacuum.

Of course the surviving $N = 1$ supersymmetry must also be eventually spontaneously broken. In this connection, we note that a third solution of $d = 11$ field equations with S^7 topology has recently been found.[13] The seven-sphere is not squashed but the F_{mnpq} components of F_{MNPQ} are now nonvanishing and provide a parallelizing torsion. This solution also admits a Higgs interpretation corresponding to nonzero vacuum expectation values for the pseudoscalars.[2,3] It also involves the group G_2 (Ref. 13) but according to D'Auria, Fre, and van Nieuwenhuizen[14] all eight supersymmetries are broken. Since our one Killing spinor η may be used to build the required totally antisymmetric torsion, namely $\bar{\eta}\Gamma_{abc}\eta$, we expect more general solutions with both squashing and torsion and this suggests an $(N = 8) \rightarrow (N = 1) \rightarrow (N = 0)$ hierarchy.

We are grateful to C. J. Isham and K. S. Stelle for helpful discussions; one of us (C.N.P.) also thanks J. P. Bourguignon and N. J. Hitchin.

---

[a] Present address: CERN, CH-1211 Geneva 23, Switzerland.

[1] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D 13, 3214 (1976); S. Deser and B. Zumino, Phys. Lett. 62B, 335 (1976).

[2] M. J. Duff and C. N. Pope, in "Supergravity 82," edited by S. Ferrara, J. G. Taylor, and P. van Nieuwenhuizen (to be published). Reported by M. J. Duff, in *Supergravity 81*, edited by S. Ferrara and J. G. Taylor (Cambridge Univ. Press, London, 1982), and by M. J. Duff and D. J. Toms, in *Unification of the Fundamental Interactions II*, edited by S. Ferrara and J. Ellis (Plenum, New York, 1982), and CERN Report No. TH. 3259, 1982 (to be published).

[3] M. J. Duff, in Proceedings of the Marcel Grossman Meeting, Shanghai, August 1982, CERN Report No. TH. 3451, 1982 (to be published).

[4] E. Cremmer and B. Julia, Nucl. Phys. B159, 141 (1979).

[5] B. de Wit and H. Nicolai, Phys. Lett. 108B, 285 (1982).

[6] J. Ellis, M. K. Gaillard, L. Maiani, and B. Zumino, in *Unification of the Fundamental Interactions*, edited by S. Ferrara, J. Ellis, and P. van Nieuwenhuizen (Plenum, New York, 1980).

[7] R. Coquereaux and M. J. Duff, to be published.

[8] P. G. O. Freund and M. A. Rubin, Phys. Lett. 97B, 233 (1980).

[9] J. P. Bourguignon and A. Karcher, Ann. Sci. Normale Sup. 11, 71 (1978).

[10] S. Kobayashi and N. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. 1.

[11] R. Gilmore, *Lie Group, Lie Algebras and Some of Their Applications* (Interscience, New York, 1974).

[12] E. Witten, Nucl. Phys. B186, 412 (1981).

[13] F. Englert, Phys. Lett. 119B, 339 (1982), and in Proceedings of the Supergravity Workshop, Trieste, September 1982 (to be published).

[14] R. D'Auria, P. Fre, and P. van Nieuwenhuizen CERN Report No. TH-3453, 1982 (to be published).

Spontaneous Supersymmetry Breaking by the Squashed Seven-Sphere

M. J. Duff,[a] B. E. W. Nilsson, and C. N. Pope[a]
Theory Group and Center for Theoretical Physics, The University of Texas at Austin,
Austin, Texas 78712

(Received 25 April 1983)

Spontaneous compactification of $d = 11$ supergravity on the squashed seven-sphere (S^7) yields a spontaneously broken phase of the $N = 8$ theory on the round S^7 with SO(8) broken to SO(5)⊗SU(2) and with $N = 1$ or $N = 0$ supersymmetry according as the squashing is left or right handed. A feature unique to Kaluza-Klein supergravity is that the massless gravitino of the $N = 1$ phase comes from a *massive* $N = 8$ supermultiplet, as do the Higgs scalars. A solution is found with squashing and torsion but only in the right-squashed $(N = 0)$ case.

PACS numbers: 04.50.+h, 04.60.+n, 11.30.Pb

The field equations of $N = 1$ supergravity in $d = 11$ dimensions admit of vacuum solutions corresponding to the product of $d = 4$ anti–de Sitter space and S^7 with its standard metric, i.e., the coset space SO(8)/SO(7). Via the Kaluza-Klein mechanism, this gives rise to an effective $d = 4$ theory with local SO(8) invariance and $N = 8$ supersymmetry.[1] The group SO(8) possesses the unique property of "triality," i.e., there are three inequivalent eight-dimensional representations with Dynkin labels (1000), (0001), (0010) to which we conventionally assign the names vector $(\underline{8}_v)$, spinor $(\underline{8}_s)$, and conjugate spinor $(\underline{8}_c)$, respectively. This triality will be crucial for what follows. In addition to the familiar massless $N = 8$ supermultiplet of spins $2, \frac{3}{2}, 1, \frac{1}{2}, 0^+, 0^-$ with SO(8) content $\{\underline{1}, \underline{8}_s, \underline{28}, \underline{56}_s, \underline{35}_v, \underline{35}_c\}$ this Kaluza-Klein theory also describes an infinite tower of massive, spin ≤ 2, supermultiplets with masses quantized in units of m, the inverse radius of S^7. They correspond to irreducible representations of OSp(4|8) but may be reducible under SO(8).[2] For example, the lightest such multiplet has the SO(8) content $\{\underline{8}_v, \underline{8}_c + \underline{56}_c, \underline{56}_v + \underline{160}_v, \underline{160}_c + \underline{224}_{vc}, \underline{112}_v, \underline{224}_{cv}\}$. The infinity of massive gravitinos is due to an infinity of spontaneously broken supersymmetries as may be seen by Fourier expansion of the $d = 11$ supersymmetry parameter[1] in harmonics on S^7. Similar remarks apply to the other $d = 11$ local symmetries of general covariance, local SO(1, 10), and gauge invariance of the three-index field. Since the SO(8) coupling constant e also depends on m via $e \sim mM_P^{-1}$ (M_P is the Planck mass), the $d = 4$ massless sector is in no sense a limiting case of the complete theory.

In a previous paper, Awada, Duff, and Pope[3] described a new "squashed" S^7 solution given by the Einstein metric on the coset space Sp(2) ⊗ Sp(1)/Sp(1) ⊗ Sp(1) where Sp(2) ≅ SO(5) and Sp(1) ≅ SU(2). Because of the topological equivalence, the $d = 4$ theory obtained from the squashed S^7 provides a spontaneously broken phase of the one obtained from the round S^7, the idea being that squashing gives nonzero vacuum expectation values (VEV's) to some of the scalar fields and hence to a Higgs and super-Higgs effect in $d = 4$.[1] SO(8) was broken to SO(5) ⊗ SU(2) and, remarkably, an $N = 1$ supersymmetry remained unbroken.

The purpose of the present paper is threefold: (1) to point out that, in contrast to the round S^7, the squashed S^7 carries a definite "handedness" and that by reversing the handedness of the $N = 1$ solution of Ref. 3, one obtains a new SO(5) ⊗ SU(2) vacuum but with $N = 0$; (2) to stress that the supersymmetry breaking mechanisms under discussion require the presence of the massive states and could not be achieved by the massless $N = 8$ supermultiplet alone; (3) to confirm the conjecture of Refs. 1 and 3 that, just as the round S^7 admits more general solutions with "torsion," so there exist S^7 solutions with both squashing *and* torsion. However, we find such solutions only in the case of the right-handed squashed $(N = 0)$ S^7 and not the left-handed squashed $(N = 1)$ S^7. In particular, we find that the field configuration suggested by Englert, Rooman, and Spindel[4] is *not* a solution of $d = 11$ supergravity. Since torsion cannot restore a broken supersymmetry, this new solution also has $N = 0$. We adopt the following conventions. The $d = 11$ space has signature $(-++\ldots)$, and $d = 11$ indices $M, N, \ldots$ will be decomposed into $d = 4$ indices

$\mu, \nu, \ldots$ and $d=7$ indices $m, n, \ldots$. The $d=11$ Dirac matrices satisfy

$$\{\Gamma_A, \Gamma_B\} = -2\eta_{AB} \tag{1}$$

and may be written

$$\Gamma_A = (\gamma_\alpha \otimes \underline{1},\ \gamma_5 \otimes \Gamma_a), \tag{2}$$

where γ_α and Γ_a are the Dirac matrices in $d=4$ and $d=7$, respectively. Spontaneous compactification works as follows. We set the fermion fields equal to zero and examine the boson field equations of the $d=11$ theory,[5]

$$R_{MN} - \tfrac{1}{2} g_{MN} R = \tfrac{1}{3}\left[F_{MPQR} F_N{}^{PQR} - \tfrac{1}{8} g_{MN} F^2\right], \tag{3}$$

$$\nabla_M F^{MNPQ} = -\tfrac{1}{576}\epsilon^{M_1 \cdots M_8 NPQ} \times F_{M_1 \cdots M_4} F_{M_5 \cdots M_8}\,. \tag{4}$$

Equation (4) is solved by the Freund-Rubin[6] choice for which all F_{MNPQ} components vanish except

$$F_{\mu\nu\rho\sigma} = 3m\epsilon_{\mu\nu\rho\sigma}, \tag{5}$$

where m is a constant. Substitution into Eq. (3) gives

$$R_{\mu\nu} = -12m^2 g_{\mu\nu} \tag{6}$$

and

$$R_{mn} = 6m^2 g_{mn}\,. \tag{7}$$

Thus the vacuum state corresponds to the product of a four-dimensional Einstein space with negative cosmological constant, which we take to be anti–de Sitter space, and a seven-dimensional Einstein space with positive cosmological constant.

As in Refs. 1 and 3, the number of unbroken supersymmetries, N, in the resulting four-dimensional theory is determined by the number of Killing spinors on the $d=7$ manifold, i.e., the number of spinors satisfying

$$\bar{D}_m \eta = (\partial_m - \tfrac{1}{4}\omega_m{}^{ab}\Gamma_{ab} - \tfrac{1}{2} m e_m{}^a \Gamma_a)\eta = 0, \tag{8}$$

where $\omega_m{}^{ab}$ and $e_m{}^a$ are the spin connection and *siebenbein* of the ground-state solution to Eq. (7). Such Killing vectors satisfy the integrability condition

$$[\bar{D}_m, \bar{D}_n]\eta = -\tfrac{1}{4} C_{mn}{}^{ab}\Gamma_{ab}\eta = 0, \tag{9}$$

where $C_{mn}{}^{ab}$ is the Weyl tensor. The subgroup of Spin(7) generated by these linear combinations of the Spin(7) generators Γ_{ab} corresponds to the holonomy group $\mathcal{H}$ of the connection of Eq. (8). Thus the maximum number of unbroken supersymmetries is equal to the number of spinors left invariant by $\mathcal{H}$.

The squashed S^7 has metric $ds^2 = ce^a \otimes ce^a$, where $c = (2\times 7^{1/2} m\lambda)^{-1}(1+8\lambda^2-2\lambda^4)^{1/2}$ is chosen to give $R = 42m^2$, and, as in Ref. 3,

$$e^0 = d\mu, \quad e^i = \tfrac{1}{2}\sin\mu\omega_i, \quad e^{\hat{i}} = \tfrac{1}{2}\lambda(\nu_i + \cos\mu\omega_i), \tag{10}$$

where $i = 1, 2, 3$ and $\hat{i} = 4, 5, 6 = \hat{1}, \hat{2}, \hat{3}$, where

$$\nu_i = \sigma_i + \Sigma_i, \quad \omega_i = \sigma_i - \Sigma_i, \tag{11}$$

and where σ_i and Σ_i are left-invariant one-forms satisfying the SU(2) algebra

$$d\sigma_1 = -\sigma_2 \wedge \sigma_3, \quad d\Sigma_1 = -\Sigma_2 \wedge \Sigma_3, \tag{12}$$

and cyclic permutations. The constant parameter λ describes the degree of distortion. The round-S^7 solution of Eq. (7) corresponds to $\lambda^2 = 1$, $C_{mn}{}^{ab} = 0$, $\mathcal{H} = \underline{1}$, and $N_{\max} = 8$. Moreover, $N = N_{\max} = 8$ because there are indeed eight solutions of Eq. (8) which we denote by η^I $(I = 1, \ldots, 8)$, where the index I is assigned to the $\underline{8}_s$ of SO(8). This is an $N=8$, SO(8) vacuum with massless states $\{\underline{1}, \underline{8}_s, \underline{28}, \underline{56}_s, \underline{35}_v, \underline{35}_c\}$. The squashed-$S^7$ solution of Eq. (7) corresponds to $\lambda^2 = \frac{1}{5}$, $C_{mn}{}^{ab} \neq 0$, $\mathcal{H} = G_2$, and $N_{\max} = 1$. Moreover, $N = N_{\max} = 1$ because there is indeed one solution of Eq. (8) which we denote by the singlet η. In contrast to the η^I of the round S^7, the components of η are actually *constant* in the orthonormal basis of Eq. (10). This yields an $N=1$, SO(5) $\otimes$ SU(2) vacuum. The massless sector consists of a 2-$\frac{3}{2}$ supermultiplet in the representation $(\underline{1}, \underline{1})$ and a 1-$\frac{1}{2}$ supermultiplet in the representations $(\underline{10}, \underline{1}) + (\underline{1}, \underline{3})$, together, possibly, with some $\frac{1}{2}$-0 supermultiplets.

The squashed S^7, however, exhibits the following subtlety. If the *left*-invariant one-forms of Eq. (12) are replaced by *right*-invariant one-forms satisfying

$$d\sigma_1 = \sigma_2 \wedge \sigma_3, \quad d\Sigma_1 = \Sigma_2 \wedge \Sigma_3, \tag{13}$$

etc., one obtains a new S^7 solution of Eq. (7), hereafter referred to as right-squashed S^7 to distinguish it from the left-squashed S^7 discussed above. Although we still have $\mathcal{H} = G_2$, there are no longer any solutions of Eq. (8), and so this vacuum has $N=0$. Because of triality, there are three inequivalent SO(5) $\otimes$ SU(2) subgroups of

SO(8). Left squashing picks out the one under which $\underline{8}_c$ goes into $(\underline{5}, \underline{1}) + (\underline{1}, \underline{3})$ and both $\underline{8}_s$ and $\underline{8}_v$ go into $(\underline{4}, \underline{2})$, while under right squashing $\underline{8}_s$ goes into $(\underline{5}, \underline{1}) + (\underline{1}, \underline{3})$ and both $\underline{8}_c$ and $\underline{8}_v$ go into $(\underline{4}, \underline{2})$. (The third subgroup may also be obtained by squashing S^7 but in an inhomogeneous manner which does not yield an Einstein metric.)

Note that by inserting a minus sign into Eq. (5) we obtain a new vacuum solution for which the criterion of unbroken supersymmetry is obtained by sending m to $-m$ in Eq. (8). On the round S^7 there are still eight solutions which we denote by $\eta^{I'}$, where the index $I' = 1, \ldots, 8$ must now be assigned to the $\underline{8}_c$ of SO(8). The SO(8) representations of the massless states are now $(\underline{1}, \underline{8}_c, \underline{28}, \underline{56}_c, \underline{35}_v, \underline{35}_s)$. All of the previous discussion on the squashed S^7 now goes through but with s and c interchanged. In this case the left-squashed sphere has $N = 0$ and the right-squashed sphere, $N = 1$. In what follows we shall retain the sign of Eq. (5).

Now we discuss the Higgs interpretation of the squashed solutions, starting with the unbroken $N = 1$ phase. First we observe that when SO(8) breaks to the appropriate SO(5) ⊗ SU(2), the massless $N = 8$ supermultiplet decomposes as follows: $\underline{1} \to (\underline{1}, \underline{1})$; $\underline{8}_s \to (\underline{4}, \underline{2})$; $\underline{28} \to (\underline{10}, \underline{1}) + (\underline{5}, \underline{3}) + (\underline{1}, \underline{3})$; $\underline{56} \to (\underline{16}, \underline{2}) + (\underline{4}, \underline{4}) + (\underline{4}, \underline{2})$; $\underline{35}_v \to (\underline{10}, \underline{3}) + (\underline{5}, \underline{1})$; $\underline{35}_c \to (\underline{14}, \underline{1}) + (\underline{5}, \underline{3}) + (\underline{1}, \underline{5}) + (\underline{1}, \underline{1})$. These states by themselves can never form $N = 1$ supermultiplets. In particular we note that *all* the eight massless gravitinos must acquire a mass. It follows that the single massless gravitino in the $N = 1$ phase must come from the *massive* sector of the $N = 8$ phase.

To proceed, we note that massless gravitinos in $d = 4$ correspond to $-\frac{7}{2}m$ modes of the Dirac operator $\not{D} \equiv \Gamma^m D_m$. That $\bar{D}_m \eta = 0$ *implies* $\not{D}\eta = -\frac{7}{2}m\eta$ is clear from Eq. (8). By consideration of the integral of $|\bar{D}_m \eta|^2$, the converse statement follows after some straightforward algebra. We therefore first calculate the Dirac operator in the squashed-S^7 metric of Ref. 3, and find

$$\not{D} = \Gamma^m \partial_m + \tfrac{3}{2}\cot\mu\,\gamma_0(1 - \gamma_5) + \tfrac{1}{4}\, i\lambda\, \gamma_0\gamma_i(1 + \gamma_5)\tau_i + (i/2\lambda)\gamma_0\gamma_i\tau_i + (3/4\lambda)\gamma_5 . \tag{14}$$

We now look for the Dirac eigenmode ψ on the squashed sphere with arbitrary λ which, in the case of $\lambda^2 = \frac{1}{5}$, is the solution η of $\bar{D}_m \eta = 0$ which gives the massless gravitino of Ref. 3. Unlike the difficult problem of solving for a generic eigenmode of $\not{D}$ the task here is very simple, by virtue of the fact that η has constant components. As may be verified by direct substitution into (14), the result is that $\psi = \eta$ for all values of λ^2, and the corresponding eigenvalue (after first rescaling the squashed-S^7 metric so that $R = 42m^2$ for all λ) is

$$-\tfrac{3}{2} \times 7^{1/2}\, m\,(1 + 2\lambda^2)(1 + 8\lambda^2 - 2\lambda^4)^{-1/2} . \tag{15}$$

Thus the eigenvalue $-\frac{7}{2}m$ mode of $\not{D}$ on the $\lambda^2 = \frac{1}{5}$ sphere has come from a $-\frac{9}{2}m$ mode on the $\lambda^2 = 1$ round sphere. Now the Dirac eigenmodes on the round S^7 can easily be calculated by projecting harmonic spinors in R^8 onto the unit sphere. One finds that the eigenvalues are $+(n + \frac{7}{2})m$ and $-(n + \frac{7}{2})m$, $n \geq 0$, the modes being in the $(n, 0, 1, 0)$ and $(n, 0, 0, 1)$ representations of SO(8), respectively. The singlet massless gravitino on the $\lambda^2 = \frac{1}{5}$ squashed S^7 has therefore come from the next-to-lowest level of Dirac eigenmodes on the round S^7, which is a $\underline{56}_c$ in the conventions of this paper. It is in fact the singlet in the decomposition $\underline{56}_c \to (\underline{10}, \underline{1}) + (\underline{10}, \underline{3}) + (\underline{5}, \underline{3}) + (\underline{1}, \underline{1})$ of SO(8) breaking down to the SO(5) ⊗ SU(2) isometry group of the left-squashed sphere.

Next we turn to the details of the Higgs mechanism. By Fourier expansion of the $g_{mn}(x^\mu, y^m)$ components of the metric in harmonics on the round S^7 one obtains the $\underline{35}_v$ scalars of the $N = 8$ massless sector as the lowest mode together with an infinite tower of massive scalars. Deformations of S^7 away from its maximally symmetric geometry correspond to nonzero VEV's for some of these scalars. In order to determine which are responsible for the symmetry breaking of this paper, we first note that in the orthonormal basis of Eq. (10), the metric deformation h_{ab} which takes the round S^7 into the squashed S^7 is of the form

$$h_{ab} \sim \mathrm{diag}(0, 0, 0, 0, 1, 1, 1) . \tag{16}$$

A straightforward calculation in the round-S^7 background shows that this is a Killing tensor,

$$\nabla_{(l} h_{mn)} = 0 . \tag{17}$$

As discussed in Ref. 1, there are 336 Killing tensors on the round S^7, transforming as a $\underline{1}$, a $\underline{35}_v$, and a $\underline{300}$ under SO(8). The $\underline{35}_v$ are associated with the massless scalars discussed above, while the $\underline{1}$ and $\underline{300}$ describe massive scalars. The singlet Killing tensor is just a constant multiple of the metric and has constant trace;

the trace of the $\underline{35}_v$ gives a $\underline{35}_v$ of eigenfunctions of the scalar Laplacian on S^7, while the $\underline{300}$ is tracefree. From (16), we see that h_{ab} has constant trace and so must be a linear combination of just the $\underline{1}$ and the $\underline{300}$. The singlet corresponds to an overall scaling of the metric and so the squashing of Eq. (10) corresponds to a nonzero VEV for the $\underline{300}$. Previously, it was suggested[1] that it might be the $\underline{35}_v$ of the massless sector which acquired a nonzero VEV. This could in fact yield a squashed S^7 with isometry group $\mathrm{SO}(5)\otimes\mathrm{SU}(2)$ but the squashing would be inhomogeneous and not a solution of Eq. (7).[7]

Finally we consider solutions with nonvanishing F_{mnpq} components of F_{MNPQ} of the kind discussed by Englert.[8] Substituting $F_{\mu\nu\rho\sigma} = 2m\epsilon_{\mu\nu\rho\sigma}$ into Eq. (4), one finds

$$\nabla_m F^{mnpq} = \tfrac{1}{6} m\, \epsilon^{npqrstu} F_{rstu} \tag{18}$$

and on the ground S^7 nontrivial solutions may be found which also solve Eq. (7). (We have rescaled $F_{\mu\nu\rho\sigma}$ so as to keep m^{-1} the radius of S^7.) Moreover the A_{mnp} field of which F_{mnpq} is the curl, $F_{mnpq} = 4\,\partial_{[m} A_{npq]}$, admits the interpretation of a parallelizing torsion. In Ref. 3, it was suggested that the constant Killing spinor η might be used to find an A_{mnp} solution on the squashed S^7. Indeed, the constant three-form $A_{abc} = \pm\frac{1}{4} \times \bar{\eta}\,\Gamma_{abc}\,\eta$ with $\bar{\eta}\eta = 1$ satisfies the Einstein equations and

$$\nabla_m F^{mnpq} = -\tfrac{1}{6} m\, \epsilon^{npqrstu} F_{rstu}\,. \tag{19}$$

However, comparison of Eqs. (18) and (19) reveals a crucial sign difference,[9] and thus the left-squashed S^7 ($N = 1$) *does not* admit solutions of this type. Had we chosen $F_{\mu\nu\rho\sigma} = -2m\epsilon_{\mu\nu\rho\sigma}$ this would be a solution, but as discussed previously this sign change is equivalent to $F_{\mu\nu\rho\sigma} = +2m\epsilon_{\mu\nu\rho\sigma}$ but with left squashing replaced by right squashing. This disagrees with the results of Ref. 4 where it is claimed that this construction may be applied to the $N = 1$ squashed S^7.

A remaining question is that of stability. The round S^7 ($\lambda^2 = 1$) and the left-squashed S^7 ($\lambda^2 = \frac{1}{5}$) are expected to be stable, because of the residual supersymmetry. Stability of both against squashing has recently been proved by Page, even though the $\lambda^2 = \frac{1}{5}$ sphere corresponds to a maximum of the potential, and $\lambda^2 = 1$ to a minimum.[10] Furthermore, we emphasize that the broken-symmetric phases of this paper cannot be extrema of the de Wit–Nicolai[11] effective potential even though their theory probably coincides with the massless sector of the $N = 8$ phase.

We are grateful to D. Page, R. Penrose, and S. Weinberg for stimulating discussions, and to J. A. Wheeler and S. Weinberg for their hospitality at the Center for Theoretical Physics and at the Theory Group. This publication was assisted by National Science Foundation Grant No. PHY 8205717 and by organized research funds of The University of Texas at Austin and the Robert A. Welch Foundation.

---

[a] On leave from the Blackett Laboratory, Imperial College, London SW7 2BZ, England.

[1] For a recent review, see M. J. Duff and C. N. Pope, in *Supersymmetry and Supergravity 82,* edited by S. Ferrara, J. G. Taylor, and P. van Nieuwenhuizen (World Scientific, Singapore, 1983).

[2] R. D'Auria and P. Fré, Phys. Lett. 121B, 141 (1983).

[3] M. A. Awada, M. J. Duff, and C. N. Pope, Phys. Rev. Lett. 50, 294 (1983).

[4] F. Englert, M. Rooman, and P. Spindel, "Supersymmetry Breaking by Torsion and the Ricci-Flat Squashed Seven-Sphere" (to be published).

[5] E. Cremmer and B. Julia, Nucl. Phys. B159, 141 (1979).

[6] P. G. O. Freund and M. A. Rubin, Phys. Lett. 97B, 233 (1980).

[7] We are very grateful to Don N. Page for bringing this to our attention.

[8] F. Englert, Phys. Lett. 119B, 339 (1982).

[9] Note that in our conventions $\Gamma_{mnpqrst} = -\epsilon_{mnpqrst}$.

[10] R. N. Page, "Classical Stability of Round and Squashed Seven-Spheres in 11-Dimensional Supergravity" (to be published).

[11] B. de Wit and H. Nicolai, Phys. Lett. 108B, 285 (1982), and Nucl. Phys. B208, 323 (1982).

ERRATA

SPONTANEOUS SUPERSYMMETRY BREAKING BY THE SQUASHED SEVEN-SPHERE. M. J. Duff, B. E. W. Nilsson, and C. N. Pope [Phys. Rev. Lett. 50, 2043 (1983)].

In our Letter we incorrectly stated that the right-handed squashed seven-sphere could be related to the left-handed sphere by changing from the left-invariant one-forms of Eq. (12) to right-invariant one-forms satisfying Eq. (13). In fact the right-squashed sphere is related to the left-squashed sphere by a reversal of orientation. This can be effected by reversing the sign of e^a $(a=0,\ldots,6)$ defined by Eq. (10). Alternatively, one can replace (10) by

$$e^0 = d\mu\,, \quad e^i = \tfrac{1}{2}\sin\mu\,\omega_i\,, \quad e^{\hat{i}} = \tfrac{1}{2}\lambda(\nu_i - \cos\mu\,\omega_i)$$

in order to obtain the right-squashed sphere. All of the discussion and conclusions in the Letter remain unchanged by this correction.

SYMMETRY BREAKING IN d = 11 SUPERGRAVITY ON THE PARALLELIZED SEVEN-SPHERE

R. D'AURIA [1], P. FRÉ [1]
Istituto di Fisica Teorica, Università di Torino, Turin, Italy
and INFN Sezione di Torino, Turin, Italy

and

P. van NIEUWENHUIZEN [1]
CERN, Geneva, Switzerland

Received 4 November 1982

In the vacuum solution proposed by Englert, in which the three-index photon of $d = 11$ supergravity plays the role of a non-vanishing torsion which parallelizes the seven-sphere, all supersymmetries are broken. A bosonic symmetry group $SO(3,2) \otimes G_2$ remains, as anticipated by Englert. These results are derived by constructing Killing vectors in the group manifold.

It has been known for a long time that the $N = 8$ (or $N = 4$) supergravity models in $d = 4$ dimensions can be most easily obtained by first considering $N = 1$ supergravity in $d = 11$ (or $d = 10$) and by considering as spacetime not a Minkovski spacetime M^{11} (or M^{10}), but rather a direct product of an M^4 with a torus T^7 (or T^6). Expanding all fields on T^7 in a Fourier series, and keeping only the zero modes, one gets after integration over the torus a massless supergravity theory in $d = 4$ with the maximal number of local supersymmetries, namely 8 [1].

Recently, there has been renewed interest in such Kaluza–Klein ideas, because other exact solutions of the $d = 11$ (or $d = 10$) field equations have been found which describe vacua with less trivial geometries. In these solutions all fermionic fields still vanish, but the photon curls [‡1] $F_{\hat{\alpha}\hat{\beta}\hat{\gamma}\hat{\delta}}$ (or $F_{\hat{\alpha}\hat{\beta}\hat{\gamma}}$) are non-vanishing and are the sources which drive the Einstein equations to these non-trivial spacetimes. In what follows, we restrict ourselves to the case $d = 11$ [‡2].

The first non-trivial solution [3] has F_{abcd} proportional to ϵ_{abcd} but all other components of $F_{\hat{\alpha}\hat{\beta}\hat{\gamma}\hat{\delta}}$ vanishing. In this sense the 3-index photon predicts the emergence of a final $d = 4$ spacetime. The energy momentum tensor satisfies $T_{ab} \sim \eta_{ab}$ and $T_{\alpha\beta} \sim \delta_{\alpha\beta}$ while $T_{a\alpha} = 0$. Thus one ends up with Einstein spaces ($R^a_b \sim \delta^a_b$ and $R^\alpha_\beta \sim \delta^\alpha_\beta$). A special case arises when one assumes that spacetime is a direct product of maximally symmetric 4- and 7- spaces

$$R^{ab}{}_{cd} = \lambda \delta^{[ab]}_{cd}, \quad R^{\alpha\beta}{}_{\gamma\delta} = \nu \delta^{[\alpha\beta]}_{\gamma\delta} . \tag{1}$$

In this case the Einstein equations show that the 4-space is an anti-deSitter space and the 7-space a deSitter space, namely a 7-sphere. This case is especially interesting because it leads to a 4-dimensional supergravity theory with eight unbroken supersymmetries [4]; in fact, its massless sector probably reproduces [4] the gauged $N = 8$ model

---

[1] Research supported by NATO under grant 236 81.

[‡1] Curved indices in $d = 11$ are denoted by $\hat{\alpha}, \hat{\beta}, \ldots$. Flat indices in $d = 4$ are denoted by $a, b, \ldots$, in $d = 7$ by $\alpha, \beta, \ldots$ and in $d = 11$ by $\hat{a}, \hat{b}$. The 11-bein is decomposed as $V^{\hat{a}} = (V^a, B^\alpha)$ and the 11-connection as $\omega^{\hat{a}\hat{b}} = (\omega^{ab}, B^{\alpha\beta}, K^{a\beta})$. The Dirac matrices in $d = 11, 7, 4$ are written as $\Gamma^{\hat{a}}, \tau^\alpha, \gamma^a$ and the τ^α are 8×8 matrices.

[‡2] The case $d = 10$ will be discussed in ref. [2].

[5]. It has been shown [6] that this vacuum has an Osp(4/8) symmetry, which is also the symmetry group one expects for the gauged $N = 8$ model in 4 dimensions [‡3].

A second solution [8] has not only F_{abcd} but also $F_{\hat{a}\hat{b}\hat{c}\hat{d}}$ non-vanishing, again in such a way that T_{ab} and $T_{\alpha\beta}$ are proportional to the 4- and 7-metrics while $T_{a\alpha} = 0$. Most interestingly, the 3-index photon $A_{\alpha\beta\gamma}$ can be interpreted as the torsion one must add to the metric 7-connection in order that the resulting 7-curvature vanishes. In that case the 7-sphere is parallelized [9].

In this letter, we will determine which symmetries the second (Englert [8]) vacuum solution has. In other words, we want to determine which supersymmetry and other bosonic transformations exist which leave the vacuum expectation values of all fields invariant up to gauge transformations. The basic idea is that Kaluza–Klein compactification of the $d = 11$ model on a space which is topologically but not geometrically equal to a 7-sphere, leads to a spontaneous broken version of the gauged $N = 8$ model [10,8]. The breaking of supersymmetry has already been studied by de Wit [11] who analyzed the commutator of the supercovariant derivatives which appear in the gravitino transformation rule. He concluded that some supersymmetries were broken; we will determine whether all supersymmetries are broken or not. To implement this programme we note that local supersymmetry transformations in x-space correspond to general fermionic coordinate transformations with transport terms (Lie derivatives) in a bigger space, namely superspace or the group-manifold. Thus we shall first "lift" the x-space vacuum fields to the bigger space. This we shall do by using the field equations of the bigger space. (We could also have used the gauge completion program for the on-shell theory [12].) It is here that we choose the group manifold approach [13] over the superspace method, because in superspace one has to impose constraints on torsions and curvatures before one can begin defining an action, whereas in the group manifold approach these constraints follow as (part of) the field equations. Having lifted all x-space fields to the bigger space, one can consider from then on this big space as superspace and study the symmetries of these fields.

The Lie derivative of a field is the difference between this field at $x^{\hat{\mu}} + \xi^{\hat{\mu}}(x)$ in the coordinate system $\bar{x}^{\hat{\mu}} = \hat{x}^{\mu} + \hat{\xi}^{\mu}(x)$ and the field at x^{μ} is the coordinate system x^{μ}. We have a physical symmetry when this difference is equal to gauge transformation; in this case $\xi^{\mu}(x)$ are the Killing vectors [14]. Since all coordinate transformations are already accounted for by $\xi^{\hat{\mu}}$, the only gauge transformations left are local SO(1,10) Lorentz rotations and Maxwell transformations. Hence, we will investigate for which fermionic vectors $\boldsymbol{q}$ the Lie derivative along $\boldsymbol{q}$ of all vacuum fields is equal to a local Lorentz rotation with parameters $(W_{\boldsymbol{q}})^{m}{}_{n}$. For example, for the 11-bein $V^{\hat{a}}(x, z, \theta)$ we must solve

$$l_{\boldsymbol{q}} V^{\hat{a}} = {}_{\underline{\boldsymbol{q}}|} \mathrm{d} V^{\hat{a}} + \mathrm{d}_{\underline{\boldsymbol{q}}|} V^{\hat{a}} = (W_{\boldsymbol{q}})^{\hat{a}}{}_{\hat{b}} V^{\hat{b}} . \tag{2}$$

Since $V^{\hat{a}}$ is known from the lifting process, one finds an equation for the Killing vector $\boldsymbol{q}$. The bosonic symmetries will be investigated in a similar way, except that the Killing vectors $\boldsymbol{k}$ will be bosonic in this case. We shall find that no fermionic $\boldsymbol{q}$ exists, i.e., all supersymmetries are broken. The bosonic symmetry will be explicitly shown to be SO(3,2) × G_2 as anticipated on general grounds in ref. [8]. We now turn to a derivation of our results.

The fields of $d = 11$ supergravity are the elfbein $V^{\hat{a}}$, the gravitino ψ and the 3-form A. They can be either considered as forms on the x-space manifold, in which case they are expanded only in $\mathrm{d}x^{\hat{\mu}}$ differentials, or as forms in superspace, in which case they have both $\mathrm{d}x^{\hat{\mu}}$ and $\mathrm{d}\theta$ parts. In the first case, the elfbeins $V^{\hat{a}}$ form a complete basis for the expansion of all other forms while in the second case a complete basis includes both $V^{\hat{a}}$ and ψ.

In the group-manifold formulation, one directly obtains the structure of a generic solution of the theory in superspace from the variational principle. This structure exhibits the property of rheonomy which amounts to saying that the complete solution in superspace is uniquely determined (up to gauge and coordinate transformations) by the solution of the field equations at x-space only.

Explicitly, for the most general solution in superspace, the curvatures 2-forms defined by [15]

$$R^{\hat{a}\hat{b}} = \mathrm{d}\omega^{\hat{a}\hat{b}} - \omega^{\hat{a}\hat{c}} \wedge \omega^{\hat{c}\hat{b}} , \quad R^{\hat{a}} = \mathrm{d}V^{\hat{a}} - \omega^{\hat{a}\hat{b}} \wedge V^{\hat{b}} - \tfrac{1}{2}\mathrm{i}\bar{\psi} \wedge \Gamma^{\hat{a}} \psi , \tag{3a,b}$$

[‡3] For a review of properties of supergravity and its gauging, see ref. [7].

$$\rho = \mathrm{d}\psi - \tfrac{1}{4}\omega^{\hat a\hat b} \wedge V^{\hat b} - \tfrac{1}{2}\mathrm{i}\bar\psi \wedge \Gamma^{\hat a}\psi\,, \quad R^{\square} = \mathrm{d}A - \tfrac{1}{2}\bar\psi \wedge \Gamma_{\hat a\hat b}\psi \wedge V^{\hat a} \wedge V^{\hat b}\,. \tag{3c,d}$$

obey the following expansion in the anholonomic basis of $(V^{\hat a}, \psi)$ [15]:

$$R^{\hat a} = 0 \equiv R^{\hat a}_{(\mathrm{sol})}\,, \quad R^{\square} = F_{\hat a_1\hat a_2\hat a_3\hat a_4} V^{\hat a_1} \wedge V^{\hat a_2} \wedge V^{\hat a_3} \wedge V^{\hat a_4} \equiv R^{\square}_{(\mathrm{sol})}\,, \tag{4a,b}$$

$$\rho = \rho_{(\mathrm{sol})} \equiv \rho_{\hat a\hat b} V^{\hat a} \wedge V^{\hat b} - \tfrac{1}{3}\mathrm{i}(\Gamma^{\hat a_1\hat a_2\hat a_3}\psi \wedge V^{\hat a_4} + \tfrac{1}{8}\Gamma^{\hat a_1\hat a_2\hat a_3\hat a_4\hat m}\psi \wedge V_{\hat m})F_{\hat a_1\ldots\hat a_4}, \tag{4c}$$

$$\begin{aligned} R^{ab} = R^{\hat a\hat b}_{(\mathrm{sol})} \equiv R^{\hat a\hat b}{}_{\hat m\hat n} V^{\hat m} \wedge V^{\hat n} + \mathrm{i}\rho_{\hat m\hat n}(\tfrac{1}{2}\Gamma^{\hat a\hat b\hat m\hat n} - \tfrac{2}{9}\Gamma^{\hat m\hat n[\hat a}\delta^{\hat b]\hat c} + 2\Gamma^{\hat a\hat b[\hat m}\delta^{\hat n]\hat c})\psi \wedge V_{\hat c} \\ + \bar\psi \wedge \Gamma^{\hat m\hat n}\psi F_{\hat m\hat n}{}^{\hat a\hat b} + \tfrac{1}{24}\bar\psi \wedge \Gamma^{\hat a\hat b\hat c_1\ldots\hat c_4}\psi F_{\hat c_1\ldots\hat c_4}\,. \end{aligned} \tag{4d}$$

The objects $F_{\hat a_1-\hat a_4}, \rho_{\hat a\hat b}, R^{\hat a\hat b}{}_{\hat m\hat n}$, which correspond to the VVVV and VV parts of the expansion and which can be identified with the supercovariant field strengths of the x-space approach [16], obey the following anholonomized propagation equations

$$\Gamma^{\hat a\hat b\hat c}\rho_{\hat b\hat c} = 0\,, \quad \mathcal{D}_{\hat m}F^{\hat m\hat c_1\ldots\hat c_3} + \tfrac{1}{96}\epsilon^{\hat c_1\ldots\hat c_3\hat a_1\ldots\hat a_8}F_{\hat a_1\ldots\hat a_4}F_{\hat a_5\ldots\hat a_8} = 0\,, \tag{5a,b}$$

$$R^{\hat a\hat m}{}_{\hat b\hat m} - \tfrac{1}{2}\delta^{\hat a}_{\hat b}R^{\hat m\hat n}{}_{\hat m\hat n} - 6F^{\hat a\hat c_1\ldots\hat c_3}F_{\hat b\hat c_1\ldots\hat c_3} + \tfrac{3}{4}\delta^{\hat a}_{\hat b}F^{\hat c_1\ldots\hat c_4}F_{\hat c_1\ldots\hat c_4} = 0\,. \tag{5c}$$

The aforementioned property of rheonomy corresponds to the fact, visible in (4), that all components of the curvatures are expressed in terms of the coefficients of the VVVV and VV parts. If one knows these objects in x-space, namely if one knows an explicit solution of the field equation (5) restricted to the x-manifold, one can insert them into (4) obtaining in this way the derivatives of all fields in θ-directions and thus one is allowed to lift the x-space fields to the whole superspace.

Note that the whole stock of eqs. (4) is obtained from the equations of motion on the group-manifold. However, (4c) and (4d) are consequences of (4a) and (4b) if one inserts them in Bianchi identities in superspace.

In the conventional superspace approach one would consider eqs. (4a) and (4c) as constraints defining the geometry and derive the rest. In the group-manifold all eqs. (4) are on the same footing and come from the variational principle.

We recall the classical solution of $d = 11$ supergravity in x-space with non-trivial torsion on the 7-sphere, using covariantly constant spinors η [10]

$$F_{a_1\ldots a_4}|_{\theta=0} = e\epsilon_{a_1\ldots a_4}\,, \quad F_{\alpha_1\ldots\alpha_4}|_{\theta=0} = a\bar\eta\tau_{\alpha_1\ldots\alpha_4}\eta\,, \tag{6a}$$

$$F_{\hat a_1\ldots\hat a_4}|_{\theta=0} = 0\,, \quad \text{otherwise}\,, \tag{6b}$$

$$R^{ab}{}_{cd}|_{\theta=0} = \lambda\delta^{[ab]}_{cd}\,, \quad R^{\alpha\beta}{}_{\gamma\delta}|_{\theta=0} = \nu\delta^{[\alpha\beta]}_{\gamma\delta}\,, \tag{6c,d}$$

$$R^{\hat a\hat b}{}_{\hat c\hat d}|_{\theta=0} = 0\,, \text{otherwise}\,, \tag{6e}$$

$$\psi_{\hat\mu}|_{\theta=0} = 0 \Rightarrow \rho_{\hat a\hat b}|_{\theta=0} = 0\,, \qquad \omega_{\hat\mu}{}^{\alpha\beta}|_{\theta=0} \equiv K_{\hat\mu}{}^{\alpha\beta}|_{\theta=0} = 0\,, \tag{6f,g}$$

where e and a are constants, one finds again for $a = 0$ the special solution of ref. [4], which has the Osp(4/8) symmetry [6]. For $a \neq 0$, one has the solution of ref. [8]. In (6a) η is a commuting 8-component Majorana (real) spinor 0-form in 7-space which satisfies $\bar\eta\eta = 1$ and

$$\mathcal{D}^{\mathrm{SO}(7)}\eta \equiv \mathrm{d}\eta - \tfrac{1}{4}B^{\alpha\beta}\tau_{\alpha\beta}\eta = bB^{\alpha}\tau_{\alpha}\eta\,. \tag{7}$$

Substituting (6) and (7) into (5), and using the identity

$$\mathbf{1} = \eta\bar\eta + \tau_{\alpha}\eta\bar\eta\tau^{\alpha}\,, \quad \bar\eta = \eta^{\mathrm{T}}\,, \tag{8}$$

one obtains:

$$\lambda = -16e^2 - 56a^2\,, \quad \nu = 4e^2 + 20a^2\,, \quad b = -\tfrac{3}{2}e\,. \tag{9}$$

The consistency of (7) relates $\mathcal{D}\mathcal{D}\eta$ to η and yields $4b^2 = \nu$, hence

$$a = \alpha e/2, \quad \alpha = \pm 1, \quad \lambda = -30e^2, \quad \nu = 9e^2 . \tag{10}$$

Calling x^a and z^a the coordinates on 4-space and 7-space, respectively, it is very easy to write down anti-deSitter 4-beins and deSitter 7-beins, and their associated metric connections satisfying (6c)–(6d) with values in (10)

$$V^a|_{\theta=0} \equiv V_0^a(x) = e^{-1}(\tfrac{2}{15})^{1/2}\mathrm{d}x^a/(1+x^2), \quad \omega^{ab}|_{\theta=0} \equiv \omega_0^{ab} = +4\mathrm{d}x^{[a}x^{b]}/(1+x^2) , \tag{11a}$$

$$B^\alpha|_{\theta=0} \equiv B_0^\alpha(z) = -e^{-1}\tfrac{2}{3}\mathrm{d}z^\alpha/(1-z^2), \quad B^{\alpha\beta}|_{\theta=0} \equiv B_0^{\alpha\beta} = -4\mathrm{d}z^{[\alpha}z^{\beta]}/(1-z^2) . \tag{11b}$$

At this point we can use the rheonomy of (4) in order to determine the first derivatives of all fields in θ on the surface $\theta = 0$ which is identified with our x-space. Splitting the indices and substituting (6) into (4) we get

$$R^a|_{\theta=0} = (\mathrm{d}V^a - \omega^{ab} \wedge V^b - \tfrac{1}{2}\mathrm{i}\bar{\psi} \wedge \gamma^a\psi - K^{\alpha\beta} \wedge B^\beta)|_{\theta=0} = 0 , \tag{12a}$$

$$R^\alpha|_{\theta=0} = (\mathrm{d}B^\alpha - B^{\alpha\beta} \wedge B^\beta - \tfrac{1}{2}\mathrm{i}\bar{\psi} \wedge \gamma_5\tau_\alpha\psi + K^{m\alpha} \wedge V_m)|_{\theta=0} = 0 , \tag{12b}$$

$$\begin{aligned} R^{ab}|_{\theta=0} &= (\mathrm{d}\omega^{ab} - \omega^{ac} \wedge \omega_\epsilon{}^b + K^{a\gamma} \wedge K^b{}_\gamma)|_{\theta=0} \\ &= 30e^2 V_0^a \wedge V_0^b + 4e\bar{\psi}_0\gamma_5\Sigma_{ab}\psi_0 - \tfrac{1}{24}\mathrm{i}\alpha e\bar{\psi}_0\tau^{\lambda_1\ldots\lambda_4}\Sigma_{ab}\psi_0\bar{\eta}\tau_{\lambda_1\ldots\lambda_4}\eta , \end{aligned} \tag{12c}$$

$$\begin{aligned} R^{\alpha\beta}|_{\theta=0} &= (\mathrm{d}B^{\alpha\beta} - B^{\alpha\gamma} \wedge B_\gamma{}^\beta + K^{m\alpha} \wedge K_m{}^\beta)|_{\theta=0} \\ &- 9e^2 B_0^\alpha \wedge B_0^\beta - \mathrm{i}e\bar{\psi}_0 \wedge \tau_{\alpha\beta}\gamma_5\psi_0 + \tfrac{1}{2}\alpha e\bar{\psi}_0 \wedge \tau^{\gamma\delta}\psi_0\bar{\eta}\tau^{\alpha\beta}\gamma\delta\eta + \tfrac{1}{2}\alpha e\bar{\psi}_0 \wedge \tau_\gamma\psi_0\bar{\eta}\tau^{\alpha\beta\gamma}\eta , \end{aligned} \tag{12d}$$

$$\begin{aligned} \rho|_{\theta=0} &= (\mathrm{d}\psi + \tfrac{1}{2}\mathrm{i}\omega^{ab} \wedge \Sigma_{ab}\psi - \tfrac{1}{4}B^{\alpha\beta}\tau_{\alpha\beta} + \tfrac{1}{2}K^{m\beta} \wedge \gamma_5\gamma_m\tau_\beta\psi)|_{\theta=0} \\ &= eB_0^\alpha\tau_\alpha\psi_0 - 2e\gamma_5\gamma_m\psi_0 \wedge V_0^m - \tfrac{1}{6}\mathrm{i}\alpha e\gamma_5\tau^{\alpha\beta\gamma}\psi_0 \wedge B_0^\delta\bar{\eta}\tau_{\alpha\beta\gamma\delta}\eta \\ &-\tfrac{1}{48}\mathrm{i}\alpha e\gamma_m\tau_{\alpha_1\ldots\alpha_4}\psi_0 \wedge V_0^m\bar{\eta}\tau^{\alpha_1\ldots\alpha_4}\eta - \tfrac{1}{4}\mathrm{i}\alpha e\gamma_5\tau_{\alpha\beta}\psi_0 \wedge B_0^\gamma\bar{\eta}\tau^{\alpha\beta}{}_\gamma\eta , \end{aligned} \tag{12e}$$

$$\begin{aligned} R^\square|_{\theta=0} &= (\mathrm{d}A + \mathrm{i}\bar{\psi} \wedge \Sigma_{ab}\psi \wedge V^a \wedge V^b - \bar{\psi}\tau_\alpha\gamma_5\gamma_b \wedge \psi \wedge B^\alpha \wedge V^b - \tfrac{1}{2}\bar{\psi} \wedge \tau_{\alpha\beta}\psi \wedge B^\alpha \wedge B^\beta)|_{\theta=0} \\ &= e\epsilon_{abcd}V_0^a \wedge V_0^b \wedge V_0^c \wedge V_0^d + \tfrac{1}{2}\alpha e\bar{\eta}\tau_{\alpha\beta\gamma\delta}\eta B_0^\alpha \wedge B_0^\beta \wedge B_0^\gamma \wedge B_0^\delta , \end{aligned} \tag{12f}$$

$$R^{\alpha\beta}|_{\theta=0} = (\mathrm{d}K^{\alpha\beta} - \omega^{ab} \wedge K^{b\beta} - B^{\beta\gamma} \wedge K^{\alpha\gamma})|_{\theta=0} = -\tfrac{1}{4}\alpha e\bar{\psi}_0 \wedge \gamma^a\tau_{\gamma\delta}\psi_0\bar{\eta}\tau^{\beta\gamma\delta}\eta . \tag{12g}$$

In the aforementioned formulae, $\psi_0 = \psi|_{\theta=0}$ denotes the value of the gravitino potential at $\theta = 0$ in superspace. In general we could write

$$\psi_0 \equiv \psi|_{\theta=0} = \psi_{\hat{\mu}}\mathrm{d}x^{\hat{\mu}} + O\mathrm{d}\theta , \tag{13}$$

where O is a 32 × 32 (x, z) dependent matrix, which, however, can always be reduced to unity by a redefinition of $\theta' = O\theta$. Since, moreover, for our solution $\psi_{\hat{\mu}}|_{t=0} = 0$, we conclude

$$\psi_0 = \mathrm{d}\theta \tag{14}$$

(We exclude the case $\theta = 0$ since in that case there would be no θ-dependence of any field.)

In the case of the solution with $a = 0$, all curvature components in (12) are constants, and transporting the right-hand sides of (12) to the left-hand sides, one recognizes the left-invariant 1-forms which satisfy an Osp(4/8) algebra in superspace [6]. These forms obviously admit 10 + 28 bosonic Killing vectors of SO(3,2) ⊗ SO(8) and 8 fermionic Killing vectors which together form an Osp(4/8) algebra.

In the present case the curvature components are no longer constants; rather they become z-dependent. Nevertheless we can perform the lifting order by order in θ obtaining all forms $V^a, B^\alpha, \omega^{ab}, B^{\alpha\beta}, \psi, K^{\alpha\beta}, A$ as polynomials in θ.

In order to decide whether the lifted potentials admit symmetry (Killing vectors), it is sufficient to know the linear part of the θ-expansion. This is so because of rheonomy: the solution is completely determined by its restriction to the $\theta = 0$ surface. However, to compute the θ-derivatives at $\theta = 0$, one needs to know the terms linear in θ of the forms. For our purposes we just need the explicit lifting of the following two potentials

$$V^a = V_0^a + \tfrac{1}{2}\mathrm{i}\bar{\theta}\gamma^a \mathrm{d}\theta + \mathrm{O}(\theta^2)\,, \tag{13a}$$

$$K^{m\beta} = \tfrac{1}{4}e(\bar{\theta}\gamma_5\gamma^a\tau_{\gamma\delta}\,\mathrm{d}\theta)(\bar{\eta}\tau^{\beta\gamma\delta}\eta) + \mathrm{O}(\theta^2)\,. \tag{13b}$$

If the solution preserves some supersymmetries, there must be at least one fermionic Killing vector leaving all the potentials invariant, in particular (13a) and (13b). Let us denote this Killing vector by $\boldsymbol{q}$. At order θ, it can be written as follows:

$$\boldsymbol{q} = H\partial/\partial\bar{\theta} + F^a\theta\,\partial/\partial x^a + F^\alpha\theta\,\partial/\partial z^\alpha + \mathrm{O}(\theta^2)\,, \tag{15}$$

where H, F^a, F^α are 32×32 x, z-dependent matrices.

The Killing equation states that the Lie derivative of the potential considered along $\boldsymbol{q}$ has to be an SO(1,10) rotation $W_{(\boldsymbol{q})}^{\hat{a}\hat{b}}$ on the same potential. Since $\boldsymbol{q}$ is fermionic, $W_{\boldsymbol{q}}^{\hat{a}\hat{b}}$ starts at order θ and therefore at $\theta = 0$ it vanishes. The invariance equations of V^a and $K^{m\beta}$ at θ are therefore

$$l_{\boldsymbol{q}}V^a|_{\theta=0} = (\boldsymbol{q}\rfloor \mathrm{d}V^a + \mathrm{d}\,\boldsymbol{q}\rfloor V^a)|_{\theta=0} = 0, \quad l_{\boldsymbol{q}}K^{m\beta}|_{\theta=0} = (\boldsymbol{q}\rfloor \mathrm{d}K^{m\beta} + \mathrm{d}\,\boldsymbol{q}\rfloor K^{m\beta})|_{\theta=0} = 0\,. \tag{16a,b}$$

From eqs. (16a,b), using (11a), (13a), (13b) and (15), one respectively obtains:

$$\tfrac{1}{2}\mathrm{i}H\gamma^a + (\tfrac{2}{15})^{1/2}e^{-1}F^a(1+x^2)^{-1} = 0\,, \quad \tfrac{1}{2}\alpha e H\gamma_5\gamma^m\tau_{\gamma\delta}(\bar{\eta}\tau^{\beta\gamma\delta}\eta) = 0\,. \tag{17a,b}$$

Eqs. (17) are just ordinary matrix equations which can be solved yielding $H = 0 = F^a$ so that $\boldsymbol{q} = 0$: the vacuum has therefore no local supersymmetry left.

Let us now turn to the bosonic symmetries of solution (6). We first solve the symmetry equation

$$l_{\boldsymbol{k}}V^{\hat{a}} = (W_{\boldsymbol{k}})^{\hat{a}}{}_{\hat{b}}V^{\hat{b}}\,, \tag{18}$$

for the elfbein $V^a = (V_0^a, B_0^\alpha)$ given in eq. (11). The most general Killing vector $\boldsymbol{k}$ of bosonic type, which satisfies (18), is a linear combination with constants coefficients

$$\boldsymbol{k} = c^a\boldsymbol{k}_a + c^{ab}\boldsymbol{k}_{ab} + c^\alpha\boldsymbol{k}_\alpha + c^{\alpha\beta}\boldsymbol{k}_{\alpha\beta}\,, \tag{19}$$

where $\{\boldsymbol{k}_a, \boldsymbol{k}_{ab}\}$ and $\{\boldsymbol{k}_\alpha, \boldsymbol{k}_{\alpha\beta}\}$ satisfy SO(2,3) and SO(8) algebras respectively

$$\begin{aligned}&[\boldsymbol{k}_{ab}, \boldsymbol{k}_{cd}] = -2\eta_{[a[c}\boldsymbol{k}_{b]d]}\,, \quad [\boldsymbol{k}_{\alpha\beta}, \boldsymbol{k}_{\gamma\delta}] = -2\eta_{[\alpha\gamma}\boldsymbol{k}_{\beta]\delta}\,, \quad [\boldsymbol{k}_{ab}, \boldsymbol{k}_c] = -\eta_{c[a}\boldsymbol{k}_{b]}\,,\\ &[\boldsymbol{k}_{\alpha\beta}, \boldsymbol{k}_\gamma] = -\eta_{\gamma[\alpha}\boldsymbol{k}_{\beta]}\,, \quad [\boldsymbol{k}_a, \boldsymbol{k}_b] = -60e^2\boldsymbol{k}_{ab}\,, \quad [\boldsymbol{k}_\alpha, \boldsymbol{k}_\beta] = 18e^2\boldsymbol{k}_{\alpha\beta}\end{aligned} \tag{20}$$

and $\eta_{ab} = (+,-,-,-)$, $\eta_{\alpha\beta} = (-,-,-,-,-,-,-)$. The explicit form of the Killing vectors $\boldsymbol{k}$ is

$$\begin{aligned}&\boldsymbol{k}_{ab} = x_{[a}\partial_{b]}\,, \quad k_{\alpha\beta} = z_{[\alpha}\partial_{\beta]}\,,\\ &\boldsymbol{k}_a = (\tfrac{15}{2})^{1/2}e[(1-x^2)\partial_a + 2x_a x\cdot\partial]\,, \quad \boldsymbol{k}_\alpha = -\tfrac{3}{2}e[(1+z^2)\partial_\alpha - 2z_\alpha z\cdot\partial]\,.\end{aligned} \tag{21}$$

In this case, the Lorentz matrices $W_{\boldsymbol{k}}^{\hat{a}\hat{b}}$ are non-vanishing elements of the SO(3,1) ⊗ SO(7) subalgebra of SO(1,10)

$$(W_{\boldsymbol{k}_c})^{ab} = -4(\tfrac{15}{2})^{1/2}e\delta_c^{[a}x^{b]}\,, \quad (W_{\boldsymbol{k}_\gamma})^{\alpha\beta} = -6e\delta_\gamma^{[\alpha}z^{\beta]}\,, \quad (W_{\boldsymbol{k}_{cd}})^{ab} = -\delta_{cd}^{[ab]}\,, \quad (W_{\boldsymbol{k}_{\gamma\delta}})^{\alpha\beta} = -\delta_{\gamma\delta}^{[\alpha\beta]}\,. \tag{22}$$

The requirement that also the 3-index photon A be invariant, which is necessary for the invariance of the total solution, translates into $l_{\boldsymbol{k}}\mathrm{A} = \mathrm{d}W_{\boldsymbol{k}}$ where $W_{\boldsymbol{k}}$ is a two-form associated with the Killing vector.

Since the Lie derivative commutes with the exterior derivative, the same condition is more easily written as

$$l_{\boldsymbol{k}}R^{\square} = 0\,, \tag{23}$$

because $R^{\square}$ is gauge invariant. It is obvious from the fact that $R^{\square}$ depends only on the z-coordinates that eq.

(23) is satisfied by all SO(2,3) Killing vectors. To investigate the 7-space symmetries we note that for a $\boldsymbol{k}$ in the SO(8) algebra we have

$$l_{\boldsymbol{k}} R^{\Box} = l_{\boldsymbol{k}} (F_{\alpha\beta\gamma\delta} B^{\alpha} \wedge B^{\beta} \wedge B^{\gamma} \wedge B^{\delta}) = 0 \,. \tag{24}$$

Recalling $l_{\boldsymbol{k}} B^{\alpha} = W_{\boldsymbol{k}}^{\alpha\beta} B^{\beta}$ and (6), (24) becomes

$$\bar{\eta}\tau^{\alpha\beta\gamma\delta} l_{\boldsymbol{k}} \eta = -2W_{\boldsymbol{k}}{}^{\alpha}{}_{\mu} \bar{\eta}\tau^{\beta\gamma\delta\mu}\eta \,. \tag{25}$$

The Lie derivative of a 0-form being equal to the differential operator $\boldsymbol{k}$ acting on a function, one evaluates $\boldsymbol{k}_{\alpha}\eta$ and $\boldsymbol{k}_{\alpha\beta}\eta$ and substitutes the result into (24).

At this point, all objects in (24) to order θ are known, except the constants $c^{\alpha\beta}, c^{\alpha}$.

To proceed, one uses the identity $I = \xi\bar{\xi} + \tau_{\alpha}\xi\bar{\xi}\tau^{\alpha}$ for *constant* ξ normalized to one, and applying this relation to η one obtains

$$\eta = X\xi + \tau_{\alpha}\xi Y^{\alpha}, \quad X = \bar{\xi}\eta, \quad Y^{\alpha} = \bar{\xi}\tau^{\alpha}\eta \,. \tag{26}$$

Since $\bar{\eta}\eta = 1$, $X^2 - Y^{\alpha} Y_{\alpha} = 1$ which shows that X and Y^{α} are coordinates on the 7-sphere. In fact, the Y^{α} are equal to $z^{\alpha}(1 - z^2)^{-1/2}$ and $X = (1 - z^2)^{-1/2}$ as one may verify, inserting (26) and (11) into (7). Replacing all η's in (25) by

$$\eta = (1 - z^2)^{-1/2} (\xi + \tau_{\alpha}\xi z^{\alpha}) \,, \tag{27}$$

one can expand (25) into powers of z^{α}. The terms cubic in z^{α} yield $c^{\alpha} = 0$, which reduces SO(8) to SO(7). The remaining terms yield:

$$c^{\alpha\beta} \bar{\xi}\tau_{\alpha\beta\gamma}\xi = 0 \,. \tag{28}$$

This is the equation which restricts the SO(7) parameters $c^{\alpha\beta}$ to those of G_2 because the 3-index constant antisymmetric tensor $\bar{\xi}\tau_{\alpha\beta\gamma}\xi$ can be identified as the G_2-invariant structure constants of the multiplication table of the octonions [17]. Thus $R^{\Box}$ is invariant only under G_2, not SO(8), and the bosonic symmetry group of the solution is SO(3,2) × G_2.

As a final remark, we mention a fact which might be related to the SU(8) symmetry of $N = 8$ supergravity in 4-dimensions and which we are studying further.

If we consider the 7-bein $B^{\alpha}_{(0)}$ and its associated metric connection $B^{\alpha\beta}_{(0)}$ as given in (11) and also the η-spinor satisfying (7) we can construct the following 1-forms on the 7-sphere:

$$\hat{B}^{\alpha} = B^{\alpha}_{0} \,, \tag{29a}$$

$$\hat{B}^{\alpha\beta} = B^{\alpha\beta}_{0} + e\bar{\eta}\tau^{\alpha\beta\gamma}\eta B_{0,\gamma} \,, \tag{29b}$$

$$\hat{B}^{\alpha\beta\gamma} = \pm(1/6\sqrt{2}) e\bar{\eta}\tau^{\alpha\beta\gamma\delta}\eta B_{0,\gamma} \,. \tag{29c}$$

These 1-forms can be interpreted as the components of an SU(8) connection. Writing for instance, in the complex 8-representation:

$$M = -\tfrac{1}{4}\hat{B}^{\alpha\beta}\tau_{\alpha\beta} + \hat{B}^{\alpha}\tau_{\alpha} + \mathrm{i}\hat{B}^{\alpha\beta\gamma}\tau_{\alpha\beta\gamma} \,, \tag{30}$$

the result is that the above defined connection is US(8) flat, namely the SU(8) Maurer–Cartan equation is satisfied

$$\mathrm{d}M + M \wedge M = 0 \,. \tag{31}$$

It is a pleasure to thank M.J. Duff and P.K. Townsend for discussions. One of us (R.D'A.) expresses his gratitude to the Dept. de Physique Théorique de l'Université de Genève for warm hospitality.

References

[1] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.
[2] M.J. Duff, P.K. Townsend and P. van Nieuwenhuizen, forthcoming paper.
[3] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233.
[4] M.J. Duff and C. Pope, reported in: M.J. Duff, Supergravity '81, eds. S. Ferrara and J.G. Taylor (Cambridge U.P., 1982); M.J. Duff, CERN preprint TH.3259 (1982).
[5] B. de Wit and H. Nicolai, Phys. Lett. B108 (1981) 285, CERN preprint TH.3291 (1982).
[6] R. D'Auria and P. Fré, University of Geneva preprint (1982).
[7] P. van Nieuwenhuizen, Phys. Rev. 68 (1981) 189.
[8] F. Englert, Phys. Lett. 119B (1982) 339; Trieste Lecture (1982).
[9] E. Cartan and J. Schouten, Proc. Kon. Akad. Wet. Amsterdam 29 (1926) 933.
[10] M.J. Duff, Shanghai Lecture, CERN preprint TH.3451 (1982).
[11] B. de Wit, Proc. Johns Hopkins workshop, Florence 1982, p. 6;
also reported by F. Englert, Trieste Lecture (1982).
[12] L. Brink, M. Gell-Mann, P. Ramond and J.H. Schwarz, Phys. Lett. 84B (1978) 336; 76B (1978) 664; 76B (1978) 417.
[13] For a review, see R. D'Auria, P. Fré and T. Regge, in: Proc. 1981 Supergravity School at Trieste, eds. S. Ferrara and J.G. Taylor;
P. van Nieuwenhuizen, Lectures at the Erice Supergravity School (1982).
[14] P. Forgács and N.S. Manton, Commun. Math. Phys. 72 (1980) 15.
[15] R. D'Auria and P. Fré, Nucl. Phys. B201 (1982) 101.
[16] E. Cremmer and S. Ferrara, Phys. Lett. 91B (1980) 61.
[17] W. Nahm, CERN preprint TH.2489 (1978), unpublished.

OCTONIONIC TORSION ON S^7 AND ENGLERT'S COMPACTIFICATION OF $d = 11$ SUPERGRAVITY ☆

Feza GÜRSEY and Chia-Hsiung TZE [1]
J.W. Gibbs Laboratory, Department of Physics, Yale University, New Haven, CT 06511, USA

Received 3 May 1983

In an octonionic realization of the Cartan–Schouten parallelism on S^7 we exhibit the algebraic nature and Hopf topology of Englert's solutions for spontaneously compactified $d = 11$ supergravity. The S^7 torsion and its dual, the index 4 gauge field are given locally by the Cayley structure constants and the associator of the imaginary octonion units, respectively.

In the current Kaluza–Klein approach to unification [1] a spontaneous compactification [2] of $d = 11$ supergravity is preferable to past compactification "by hand" [3]. Thus the Freund–Rubin–Englert (FRE) solution [4,5] compactifies $d = 11$ spacetime into a Riemann product of two Einstein spaces, a 4-dimensional anti-de Sitter universe and S^7, the 7-sphere of the unit octonions.

In the Bose sector of $d = 11$ supergravity [3], the Freund–Rubin solution [4] is such that the index 4 fully antisymmetric gauge field strength F_{MNPQ} with $M, N, P, Q = 0, 1, 2, ..., 10$ only takes non-zero expectation values in its 4-spacetime components $F_{\mu\nu\rho\sigma} = f g_4^{1/2} \epsilon_{\mu\nu\rho\sigma}$ with $\mu, \nu, \rho, \sigma = 0, 1, 2, 3$, $f = \text{cst}$, g_4 is the determinant of the spacetime metric and $\epsilon_{\mu\nu\rho\sigma}$ is the Levi-Civita dualizing symbol. Englert's solutions [5] are such that $F_{MNPQ} \neq 0$ not only for the spacetime components assuming the Freund–Rubin form, but also for the internal space components $m, n, p, q = 4, 5, ..., 10$. Apart from a constant multiplicative factor these F_{mnpq} can be identified with the skew 4 index tensor dual to the fully antisymmetric 3 index Cartan–Schouten torsion on S^7. While the Freund–Rubin solution corresponds to one of three Cartan affine metric geometries on S^7, namely to the torsionless riemannian geometry, the two Englert solutions, differing by an overall relative sign, realize the remaining (left and right) curvature-less but torsion-full Cartan geometries of absolute parallelism (AP for short) on S^7 [6]. Physically these FRE solutions provide a geometric Higgs mechanism [7] leaving a residual $SO(3,2) \otimes G_2$ bosonic symmetry. They underscore the long suspected geometrical nature of the index 4 gauge field strength of SO(8) supergravity. Moreover through the relevance of octonions has long been apparent in supergravity theories. it can be made most conspicuous by performing an octonionic analysis of Englert's results.

While they attributed the cause of the AP property on S^7 to the non-associativity of the octonions, neither ref. [6] nor latter works [8,10] [‡1] known to us have uncovered the singular connection which must exist between the *local* S^7 differential geometry of absolute parallelism and the alternative Cayley algebra [11]. In ref. [5] while the torsion coefficients of Lie group manifolds, the only other nontrivial compact spaces with the AP property besides S^7, are showed to be given simply by Lie's structure constants, their counterparts on S^7 were left as abstract entities. Englert's work [4] inherits this lack of explicitness. We will

---

☆ Research supported in part by the US Department of Energy under contract No. DE-AC02-76ER03075.

[1] A.P. Sloan Foundation Fellow.

[‡1] See Wolf [9] for the most up-to-date analysis of absolute parallelism.

present an explicit algebraic octonionic solution to the Cartan–Schouten–Englert equations for the S^7 torsion and its differential invariants. Only results are given, details will be part of a larger work [12]. We begin by grafting the natural algebraic octonionic structure onto the S^7 geometry.

Briefly stated, absolute parallelism on S^7 [6] is due to the existence [13] on that manifold of 7 orthonormal global Killing vector fields or equivalently to that of a Cayley multiplication $e_i e_j = -\delta_{ij} + \psi_{ijk} e_k$ on S^7. The e_i, $i = 1, 2, \ldots, 7$ are the 7 octonionic imaginary units with $\{e_i, e_j\} = -2\delta_{ij}$ and the ψ_{ijk} are the fully antisymmetric structure constants [14]. This existence of a H-multiplication [15] of R^7 × R^7 → R^7 of the e is, in turn, equivalent [16] to the reduction of the group SO(7) of the frame bundle on S^7 to its maximal subgroup G$_2$, the automorphism group of the imaginary octonion units. In fact as a coset space S^7 $\approx$ SO(8)/SO(7) $\approx$ SO(7)/G$_2$.

Let us consider on S^7, taken as a topological and geometrical 7-sphere, a coordinate patch x^α ($\alpha = 1, 2, \ldots, 7$) and its metric $\mathrm{d}s^2 = g_{\alpha\beta}\,\mathrm{d}x^\alpha \mathrm{d}x^\beta = \delta_{ij} h^i_\alpha\binom{\mathrm{L}}{\mathrm{R}} \times h^j_\beta\binom{\mathrm{L}}{\mathrm{R}}\,\mathrm{d}x^\alpha \mathrm{d}x^\beta$. The $h^i_\alpha\binom{\mathrm{L}}{\mathrm{R}}$, $i = 1, 2, \ldots, 7$ are left (L) and right (R) septads or sieben-bein, respectively. We then define the octonionic-valued covariant 7-legs $\widetilde{E}_\alpha\binom{\mathrm{L}}{\mathrm{R}} = e_i h^i_\alpha\binom{\mathrm{L}}{\mathrm{R}}$; they obey the relations

$$\{\widetilde{E}_\alpha\tbinom{\mathrm{L}}{\mathrm{R}}, \widetilde{E}_\beta\tbinom{\mathrm{L}}{\mathrm{R}}\} = -2g_{\alpha\beta}\,, \tag{1}$$

$$[\widetilde{E}_\alpha\tbinom{\mathrm{L}}{\mathrm{R}}, \widetilde{E}_\beta\tbinom{\mathrm{L}}{\mathrm{R}}] = 2S_{\alpha\beta\gamma}\widetilde{E}^\gamma\tbinom{\mathrm{L}}{\mathrm{R}}\,, \tag{2}$$

where

$$\begin{aligned} S_{\alpha\beta\gamma}\tbinom{\mathrm{L}}{\mathrm{R}} &\equiv \psi_{ijk} h^i_\alpha\tbinom{\mathrm{L}}{\mathrm{R}} h^j_\beta\tbinom{\mathrm{L}}{\mathrm{R}} h^k_\gamma\tbinom{\mathrm{L}}{\mathrm{R}} \\ &\equiv \mathrm{Sc}\,\{\widetilde{E}_\alpha\tbinom{\mathrm{L}}{\mathrm{R}}\widetilde{E}_\beta\tbinom{\mathrm{L}}{\mathrm{R}}\widetilde{E}_\gamma\tbinom{\mathrm{L}}{\mathrm{R}}\}\,, \end{aligned} \tag{3}$$

is a fully skew-symmetric index 3 tensor. Sc denotes the scalar part of an octonion $a \in \Omega$ where Ω is the division ring of Cayley numbers. $a = a_i e_i$, $a_i \in \mathrm{R}$, the real number field, $e_0 = 1$ is the eigth octonion unit. $\bar{a} = (a_0, -a) = a_i \bar{e}_i$. So $\mathrm{Sc}(a) = \frac{1}{2}(a + \bar{a})$, and the vectorial part $\mathrm{Vec}(a) = \frac{1}{2}(a - \bar{a})$.

As in the associative Lie group case [17], we define absolute parallelism on S^7 through Cartan's first structural equations [8,13]

$$\nabla_\alpha \widetilde{E}_\beta\tbinom{\mathrm{L}}{\mathrm{R}} = \mp T^\gamma_{\alpha\beta}\widetilde{E}_\gamma\tbinom{\mathrm{L}}{\mathrm{R}}\,, \tag{4}$$

where $\nabla_\alpha \widetilde{E}_\beta = \partial_\alpha \widetilde{E}_\beta - \mathring{\Gamma}^\gamma_{\alpha\beta}\widetilde{E}_\gamma$ is the covariant derivative w.r.t. the symmetric riemannian connection

$$\mathring{\Gamma}^\gamma_{\alpha\beta} = \left\{{\gamma \atop \alpha\beta}\right\} \equiv \tfrac{1}{2}(\Gamma^\gamma_{\alpha\beta} + \Gamma^\gamma_{\beta\alpha})\,,$$

$T^\gamma_{\alpha\beta} = \frac{1}{2}(\Gamma^\gamma_{\alpha\beta} - \Gamma^\gamma_{\beta\alpha})$ is the torsion tensor. $\Gamma^\gamma_{\alpha\beta}$ is the asymmetric affine connection [12]. By eq. (4) the $\widetilde{E}_\alpha\binom{\mathrm{L}}{\mathrm{R}}$ are parallel transported w.r.t. the left and right connection $\Gamma^\gamma_{\alpha\beta}\binom{\mathrm{L}}{\mathrm{R}} = (\mathring{\Gamma}^\gamma_{\alpha\beta} \pm T^\gamma_{\alpha\beta})$. With the metric $g_{\alpha\beta}$ being covariantly constant w.r.t. all the above three connections, the torsion tensor is then antisymmetric in all *three* indices. And $T_{\alpha\beta\gamma}(\mathrm{L}) = -T_{\alpha\beta\gamma}(\mathrm{R})$. The remaining defining equations are the second structural equations:

$$[\nabla_\alpha, \nabla_\beta]\,\widetilde{E}_\gamma\tbinom{\mathrm{L}}{\mathrm{R}} = R_{\alpha\beta\gamma\sigma}\tbinom{\mathrm{L}}{\mathrm{R}}\widetilde{E}^\sigma\tbinom{\mathrm{L}}{\mathrm{R}}\,, \tag{5}$$

where the non-integrable riemannian curvature w.r.t. $\left\{{\gamma \atop \alpha\beta}\right\}$ is

$$R_{\alpha\beta\gamma\sigma} = T_{\rho\alpha\beta}T^\rho_{\gamma\sigma} - T_{\rho[\alpha\beta}T^\rho_{\gamma]\sigma}\,, \tag{6}$$

$R_{\alpha\beta\gamma\sigma}(\mathrm{L}) = R_{\alpha\beta\gamma\sigma}(\mathrm{R})$ as it depends solely on the symmetric connection.

In seeking an explicit form for $T_{\alpha\beta\gamma}$, we recall [17] that for a rank r compact Lie group G $T_{\alpha\beta\gamma} = \pm\frac{1}{2}f_{ijk} \times w^i_\alpha w^j_\beta w^k_\gamma$ where f_{ijk} are the Lie structure constants and the $w^i_\alpha\binom{\mathrm{L}}{\mathrm{R}}$ are the components of the r Killing vector fields on G relative to a given local coordinate basis. Now the group torsion is covariantly constant: $\nabla_\mu T_{\alpha\beta\gamma} = T_{\rho[\alpha\beta}T_{\gamma}{}^{\rho}{}_{]\mu} = 0$. Its right hand side, being nothing but the Jacobi identity, vanishes by virtue of the *associativity* of Lie algebras. It follows that the w^i_α obey the well known Maurer–Cartan integrability conditions. The absolute parallelism on G is thus called integrable. We observe that $\nabla_\mu T_{\alpha\beta\gamma} = 0$ is but another guise for the Poincaré lemma $\mathrm{d}^2 = 0$ [18], d denoting exterior differentiation. That lemma, an axiom carried over from euclidean geometry, implies the flatness of the tangent space of a riemannian manifold. In a non-associative analogy to Lie groups, we posit the S^7 torsion to be locally of the form

$$T_{\alpha\beta\gamma} = R_0^{-1}S_{\alpha\beta\gamma}\tbinom{\mathrm{L}}{\mathrm{R}} = \pm R_0^{-1}\psi_{ijk}h^i_\alpha\tbinom{\mathrm{L}}{\mathrm{R}}h^j_\beta\tbinom{\mathrm{L}}{\mathrm{R}}h^k_\gamma\tbinom{\mathrm{L}}{\mathrm{R}}\,, \tag{7}$$

where R_0 is the S^7 radius and ψ_{ijk} are the octonionic structure constants. We proceed to show that the ansatz eq. (7) does in fact satisfy all the Cartan–Schouten–Englert equations for $T_{\alpha\beta\gamma}$. The latter equations can be said [19] to define a *non-integrable* parallelism on S^7. In the sequel we will need some key identities obeyed by the ψ_{ijk}:

Due to the alternativity of the Cayley algebra

[8,11], the associator $(a,b,c) = (ab)c - a(bc)$, is fully antisymmetric in any three octonions a, b and c; e.g. $(a,b,b) = 0$. From $(e_i, e_j, e_j) = 0$ (where we sum over j) we obtain

$$\psi_{irs}\psi_{jrs} = 6\delta_{ij} . \tag{8}$$

Similarly using the Moufang identity $(xa)(bx) = x(ab)x, x \in \Omega$, we get

$$\psi_{ris}\psi_{sjt}\psi_{tkr} = 3\psi_{ijk} . \tag{9}$$

For the basic associator of any three imaginary units we have

$$(e_i, e_j, e_k) = 2\varphi_{ijkr}e_r \tag{10}$$

$$= 2\psi^{l}_{[ij}\psi^{l}_{k]r}e_r , \tag{11}$$

where the fully antisymmetric 4 index entity φ_{ijkr} is also given through our *fundamental* identity

$$\varphi_{ijkr} = (1/3!)\,\epsilon_{ijkrlmn}\psi_{lmn} \tag{12a}$$

$$= (\delta_{jk}\delta_{ir} - \delta_{ik}\delta_{jr}) + \psi_{ijs}\psi_{krs} . \tag{12b}$$

Our other key observation is the duality connection:

$$\psi_{ijk} = -(1/4!)\,\epsilon_{ijklmnr}\varphi_{lmnr}$$

$$= -(1/4!)\,\epsilon_{ijklmnr}\psi_{lms}\psi_{nrs} , \tag{13}$$

which can be read off from the following variant of Cayley's multiplication table [14]:

$$\psi_{ijk}\begin{cases} 1 & 2 & 4 & 3 & 6 & 5 & 7 \\ 2 & 4 & 3 & 6 & 5 & 7 & 1 \\ 3 & 6 & 5 & 7 & 1 & 2 & 4 \end{cases} \tag{14a}$$

and

$$\varphi_{lmnr}\begin{cases} 4 & 3 & 6 & 5 & 7 & 1 & 2 \\ 5 & 7 & 1 & 2 & 4 & 3 & 6 \\ 6 & 5 & 7 & 1 & 2 & 4 & 3 \\ 7 & 1 & 2 & 4 & 3 & 6 & 5 \end{cases} \tag{14b}$$

The above table reads as follows: each line is a cyclic permutation of the first line, the columns of the first three lines give the structure constants or half the commutators of the octonionic imaginary units e_i, e.g. $e_1e_2 = e_3, e_2e_4 = e_6$ etc... while the columns of the remaining four lines give half the associator of the e_i e.g. $\frac{1}{2}(e_3, e_7, e_5) = -e_1, \frac{1}{2}(e_4, e_5, e_6) = -e_7$, etc... hence they determine the φ_{ijkr} e.g. if $\psi_{123} = 1, \varphi_{4567} = -1$ etc...

Indeed the algebraic duality eq. (13) constitutes the kernel of our solution for the Cartan–Schouten torsion on S^7, as we now show. Using the form eq. (7) for the torsion we easily verify that the two Cartan–Schouten equations

$$T^{\lambda\sigma}_{\alpha}T_{\lambda\sigma\beta} = 6R_0^{-2}g_{\alpha\beta} , \tag{15a}$$

$$T^{\alpha}_{\lambda\sigma}T^{\gamma}_{\alpha\beta}T^{\lambda}_{\gamma\rho} = 3R_0^{-2}T_{\beta\sigma\rho} , \tag{15b}$$

are mere septad-dressed i.e. covariant forms of the algebraic identities eqs. (8) and (9) respectively. The latter coupled with eq. (4) reproduce the riemannian geometry of S^7 as a locally symmetric Einstein space with Riemann, Ricci and scalar curvatures $R_{\alpha\beta\rho\sigma} = R_0^{-2}(g_{\alpha\rho}g_{\beta\sigma} - g_{\alpha\sigma}g_{\beta\rho}); R_{\alpha\beta} = 6R_0^{-2}g_{\alpha\beta}$ and $R = 42R_0^{-2}$, respectively. While $\nabla_\alpha T_{\alpha\beta\gamma} = 0$ holds,

$$\nabla_\rho T_{\alpha\beta\gamma} = T_{\lambda\rho[\alpha}T^{\lambda}_{\beta\gamma]} \equiv T_{[\alpha\beta\gamma,\rho]} , \tag{16}$$

is non-zero since, unlike the Lie constants f_{ijk}, the ψ_{ijk} do *not* obey the Jacobi identity. In its place, we have the relation

$$\psi_{ijk}\varphi_{ijkr}e_r = 0 , \tag{17a}$$

or

$$\psi_{ijk}\psi^{l}_{[ij}\psi^{l}_{k]r} = 0 , \tag{17b}$$

which generalizes Jacobi's identity and hence Poincaré's lemma to octonions [12].

Similarly using eq. (10) we obtain Englert's equation [5];

$$\Delta T_{\alpha\beta\gamma} = -18R_0^{-2}T_{\alpha\beta\gamma} , \tag{18}$$

where Δ is the Laplace–Beltrami operator on S^7. Taking the dual of both sides of eq. (16) we have

$$g_7^{-1/2}\,\epsilon^{\lambda\mu\nu\rho\alpha\beta\gamma}\nabla_\rho T_{\alpha\beta\gamma} = 3g_7^{-1/2}\,\epsilon^{\lambda\mu\nu\rho\alpha\beta\gamma}T^{\sigma}_{\beta\rho}T_{\beta\gamma\sigma} , \tag{19}$$

whose rhs is just the covariant form of the identity eq. (13) between the dual of the "Cayley curvature" φ_{ijkl} and the ψ_{ijk}. It is a striking fact that with $\nabla_\rho T_{\alpha\beta\gamma}$ being non-zero and fully antisymmetric in all four indices, the relation eq. (6) or using eq. (16), $R_{\alpha\beta\gamma\sigma} = \nabla_\gamma T_{\beta\alpha\sigma} + T_{\rho\alpha\beta}T^{\rho}_{\gamma\sigma}$ between torsion and riemannian curvature of a locally symmetric space can only be satisfied in *seven* dimensions [13] and moreover that it can be fulfilled to simply by the G_2 invariant identity eq. (12) dressed with the legs h^i_α! The Riemann curvature for S^7 is thus given by the sum of a "Cayley curvature" and a "Lie curvature"; the first being zero for the associative case of Lie groups.

Eq. (19) solves for Englert's 7-dimensional covariant equation for the 4 index gauge field $F^{\alpha\beta\gamma\rho}$

$$\nabla_\alpha F^{\alpha\beta\gamma\rho} = k g_7^{-1/2} \epsilon^{\beta\gamma\rho\sigma\zeta\mu\nu} F_{\sigma\zeta\mu\nu} \tag{20}$$

(k is a constant), with the identification $F_{\alpha\beta\gamma\rho} = \lambda T_{[\alpha\beta\gamma,\rho]}$, λ = a constant [4]. From eqs. (16) and (11) for the basic associator tensor φ_{ijkl} and in terms of the legs h^i_α we have explicitly

$$F_{\alpha\beta\gamma\rho} = \lambda R_0^{-1} \varphi_{ijkr} h^i_\alpha h^j_\beta h^k_\gamma h^r_\rho \,.$$

Thus from our octonionic vantage point the non-vanishing nature of the $F_{\alpha\beta\gamma\rho}$, and hence of the ensuing geometric Higgs mechanism, is traced directly to the alternativity and non-associativity of the Cayley numbers, to their non-zero associator! It suffices to add that with the help of another identity [12]

$$\varphi^{ilmn}\varphi_{jlmn} = 24\delta^i_j \,, \tag{21}$$

we recover the remaining compactification formulae of Englert [5] in purely algebraic forms.

In closing a few remarks are in order. Our assumed expression for $T_{\alpha\beta\gamma}$, eq. (8), which yields an algebraic solution to the Cartan–Schouten S^7 torsion [6], requires a deeper explanation. The latter is in our view tied to the existence of an *integrable* absolute parallelism on S^7. We model the latter after the quaternionic integrable parallelism on S^3 [20,12]. Again only pertinent results are stated; their derivation involves a judicious manipulation of Moufang's along with new G_2-invariant octonionic identities [12].

By way of a Cayley projection map on R^8, the 7-sphere of radius R is defined by the equation $Y\bar{Y} = R_0^2$, the octonion $Y = R_0 V$ with $V\bar{V} = 1$ and

$$V = \frac{1 + \xi/2R_0}{1 - \xi/2R_0} \equiv 2L - 1 \,, \tag{22}$$

where $L \equiv (1 - \xi/2R_0)^{-1}$, $\xi = e\cdot\boldsymbol{\xi}$; ξ^α ($\alpha = 1, 2, ..., 7$) are the stereographic coordinates. We introduce the left and right integrable 7-legs $E_\alpha(\mathrm{L})$ and $E_\alpha(\mathrm{R})$ in the forms of

$$\omega_\mathrm{L} \equiv E_\alpha(\mathrm{L})\,\mathrm{d}\xi^\alpha = R_0 \bar{V}\,\mathrm{d}V = \bar{L} e_\alpha L\,\mathrm{d}\xi^\alpha \,,$$

$$\omega_\mathrm{R} \equiv E_\alpha(\mathrm{R})\,\mathrm{d}\xi^\alpha = R_0\,\mathrm{d}V\bar{V} = L e_\alpha \bar{L}\,\mathrm{d}\xi^\alpha \,, \tag{23}$$

giving the conformally flat line element of S^7:

$$\mathrm{d}s^2 = |\omega_{\mathrm{L,R}}|^2 = (1 + \xi^\alpha\xi_\alpha/4R_0^2)^{-2}\,\mathrm{d}\xi^\alpha\,\mathrm{d}\xi_\alpha$$

$$\equiv \phi^2\,\mathrm{d}\xi^\alpha\,\mathrm{d}\xi_\alpha \,,$$

with the Riemann metric $g_{\alpha\beta} = \phi^2\delta_{\alpha\beta}$. We check that the orthonormal condition $E^a_\alpha(^\mathrm{L}_\mathrm{R})\,E^a_\beta(^\mathrm{L}_\mathrm{R}) = g_{\alpha\beta}$ as well as the Killing property $\nabla_\alpha E_\beta + \nabla_\beta E_\alpha = 0$ hold [13]. We obtain the explicit expression:

$$E^a_\alpha(^\mathrm{L}_\mathrm{R}) = \phi^2\{[1 - (1/4R_0^2)\,\xi^\rho\xi^\rho]\,\delta_{\alpha a} + (1/2R_0^2)\xi^\alpha\xi^a$$

$$\pm (1/R_0)\,\psi_{\alpha\rho a}\,\xi^\rho\} \,, \tag{24}$$

which are in exact correspondence with the quaternionic legs on S [20,12].

By *integrable* absolute parallelism we further understand the following prescription for the parallel left (right) transportation of the unit octonion vector $\bar{V}(V)$

$$D^\mathrm{L}_\alpha \bar{V} \equiv (\partial_\alpha + R_0^{-1} E_\alpha(\mathrm{L}))\,\bar{V} = 0 \,,$$

$$D^\mathrm{R}_\alpha V \equiv (\partial_\alpha - R_0^{-1} E_\alpha(\mathrm{R}))\,V = 0 \,, \tag{25}$$

i.e. $\bar{V}(V)$ is covariantly constant. Here the $D^{\mathrm{L,R}}_\alpha$ are new octonionic covariant affine derivatives and hence do *not* admit any ordinary matrix representation. After some (non-associative) algebra, we find

$$\partial_\alpha E_\beta(\mathrm{R}) - \partial_\beta E_\alpha(\mathrm{R}) = 2\phi R_0^{-1}\psi_{\alpha\beta\gamma} E_\gamma(\mathrm{R})$$

$$- 2R_0^{-1} L^2 \{L[L^{-1}, E_\alpha(\mathrm{R}), E_\beta(\mathrm{R})]\} L^{-2} \,. \tag{26}$$

We omit the corresponding equation for $E_\alpha(L)$. Clearly eq. (26) is the non-associative extension of the Maurer–Cartan integrability conditions for Lie groups. In the latter case [17,12], the rhs of eq. (26) would only have the first term. It is precisely the presence of the additional associator term on the rhs of eq. (26) which makes the integrable parallelism on S^7 much more subtle! It seems therefore all the more remarkable that by mimicking the Lie group torsional geometry, albeit with the Cayley structure constants, we succeeded in satisfying all the Cartan–Schouten–Englert (CSE) equations. What then is the connection, if any, between the legs $\widetilde{E}_\alpha(^\mathrm{L}_\mathrm{R})$ and $E_\alpha(^\mathrm{L}_\mathrm{R})$ of the just defined global parallelism which does *not* satisfy the CSE equations? This question is currently under active investigation. Specifically, we are seeking an explicit local SO(7) transformation relating these two sets of legs.

Finally we comment briefly on the topology of Englert's solutions. We recall the simple case of the 1-sphere, a circle of radius R_0, its leg $R_0\,\mathrm{d}\theta$ maps S′ onto S′ with winding number ±1 depending on orientation i.e. left or right translation. The Chern–Simons winding index is given by the volume of the unit circle $(2\pi)^{-1} \oint \mathrm{d}\theta = \pm 1$. Thus the legs are associated with the 1-Dirac U(1) monopole bundle, the complex Hopf fibering of S^3 by $S^2 = CP(1)$ with fiber S′ [21]. Similarly on S^3 the quaternionic valued legs E_α are $S^3 \to S^3$ maps of winding number ±1. The latter is given by integration over S^3 of the harmonic volume 3-form, the parallelizable torsion $T_{\alpha\beta\gamma}$ as

$$(24\pi^2)^{-1} R_0 \oint_{S^3} \epsilon^{\alpha\beta\gamma} T_{\alpha\beta\gamma}\, \mathrm{d}^3x = \pm 1 , \tag{27}$$

normalized to yield the volume of an unit 3-sphere. Eq. (27) is the familiar winding number for Skyrme's S^3 hedgehog or the 1-BPST instanton [21,22]. The legs E_α realize the canonical connections of the quaternionic Hopf fibering of S^7 over $S^4 = HP(1)$ with S^3 as fiber and Hopf invariant [1]. In the octonionic case, the $\widetilde{E}_\alpha$ and hence the torsion $T_{\alpha\beta\gamma}$ i.e. the index 3 gauge potential $A_{\alpha\beta\gamma}$ and $F_{\alpha\beta\gamma\rho}$ are $S^7 \to S^7$ mappings of winding degree ±1. The associated invariant is the volume of the unit 7-sphere S^7 given by

$$(252\pi^4)^{-1} R_0^5 \oint_{S^7} \epsilon^{\alpha\beta\gamma\rho\sigma\mu\nu} F_{\rho\sigma\mu\nu} A_{\alpha\beta\gamma}\, \mathrm{d}^7x = \pm 1 , \tag{28}$$

whose generalized Chern–Simons integrand is thus a harmonic 7-form [23] [‡2]. Physically the quantity eq. (28) contributes to the cosmological terms in Englert's solutions [4]. The above parallels induce us to identify Englert's solutions with the last Hopf fibering [24], namely that of S^{15} over $S^8 = \Omega P(1)$, the octonionic projective line with as fiber S^7 and with Hopf invariant [1]. By a celebrated theorem of Adams (and Atiyah) [15], if there exists a Hopf map: $S^{2n-1} \to S^n$ of Hopf invariant 1, then n must equal 2, 4 or 8. The existence of these "essential" maps including that of $S^0 \approx Z_2 \to S'$ with fiber S′ [25] are uniquely linked to the existence of only 4 division algebras over the real numbers. In more down-to-earth terms, the above observations suggest that physically Englert's solutions are the 7-dimensional analogs of the kink [25], the U(1)-monopole and SU(2)-instanton [21]. The full implications of the exceptional topology of the 4 index gauge field sketched here, its associated Morse theory and Jacobi fields for the physics of Kaluza–Klein supergravity surely merits a detailed study.

From the above perspective, while Englert's solutions occupy a rather unique geometric position they should not deter us from seeking solutions with winding number higher than 1 as have been done for the monopole and instanton. We have been searching for these solutions by way of an octonionic extension of Fueter's hypercomplex analyticity [26]. It is well known that $O(4) \approx S^3 \otimes S^3$, the norm group of the quaternions, underlies the rich duality structure of euclidean SU(2) Yang–Mills fields and Einstein gravity in four dimensions [21]. $SO(8) \approx S^7 \otimes S^7 \otimes G_2$, the norm group of the octonions, with its remarkable Triality principle, turns out to be endowed with a more exotic duality in 8 dimensions. A residual aspect of this duality rooted in G_2 is already visible thanks to Englert [4] in 7 dimensions and in an explicit algebraic setting here. Moreover in our view the non-associative Maurer–Cartan relations eq. (26) open the way to the formulation of a possibly novel non-riemannian and non-desarguean geometry of parallelism in 7 and 8 dimensions. In a larger companion paper [12] we will discuss these topics along with providing details on the results reported here.

One of us (F.G.) wishes to warmly thank Professor Michael Atiyah for inspiring remarks on the differential geometric implications of non-associativity, Professor F. Englert for informing him that Ms. M. Rooman of the Université Libre the Bruxelles has also been engaged in an octonionic analysis of his solutions. We are also grateful to our Yale colleagues, particularly to Howard Garland and Greg Zuckerman for enlightening discussions on topics related to this work.

[‡2] The 7-form in eq. (28) is therefore a generator of $H^7(S^7)$, the seventh De Rham cohomology group of S^7 whose dimension is one. So the seventh Betti number of S^7 is one, the number of harmonic 7-form.

References

[1] See for instance: E. Witten, Nucl. Phys. B186 (1981) 412;
A. Salam and J. Strathdee, Ann. Phys. (NY) 141 (1982) 316, and references therein.
[2] M.J. Duff, in: Supergravity 81, eds. S. Ferrara and J.G. Taylor (Cambridge U.P., London, 1982);
M.J. Duff and D.J. Toms, in: Unification of the fundamental interactions, eds. J. Ellis and S. Ferrara (Plenum, New York, 1982).
[3] F. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.
[4] P.G.O. Freund and M. Rubin, Phys. Lett. 97B (1980) 233.
[5] F. Englert, Phys. Lett. 119B (1982) 339.
[6] E. Cartan and J.A. Schouten, Proc. K. Akad. Wet. Amsterdam 29 (1926) 803, 933;
see also A. Einstein, Berl. Ber. (1928).
[7] M.J. Duff, Supergravity, the 7-sphere and spontaneous symmetry breaking, CERN-TH-3451 (1982);
R. D'Auria, P. Fré and P. van Nieuwenhuizen, Phys. Lett. 122B (1983) 225.
[8] R. Hermann, Spinor, Clifford and Cayley algebras, Interdisciplinary Mathematics (Robert Hermann, Brookline, Ms., 1974), Vol. VIII, Ch. 5.
[9] J. Wolf, J. Diff. Geom. 4 (1972) 317; 7 (1972) 19.
[10] F. van der Blij, History of the octaves, in: Simon Stevin, Wis- en Natuurkundig Tijdschrift 34e Jaargang, Aflevering III (1961) 106;
see in particular: F. Vaney, Le parallèlisme absolu dans les espaces elliptiques réels à 3 at à 7 dimensions et le principle de trialité, Thèse Paris 1929;
E.A. Weiss, Math. Z. 44 (1938) 5808.
[11] R.D. Schafer, An introduction to non-associative algebras (Academic Press, New York, 1966).
[12] F. Gürsey and C.-H. Tze, Yale preprint in preparation.
[13] J.E. D'Atri and N.K. Nickerson, J. Diff. Geom. 2 (1968) 393.
[14] M. Günaydin and F. Gürsey, J. Math. Phys. 14 (1973) 1651.
[15] D. Husemoller, Fibre bundles (McGraw Hill, New York, 1966).
[16] A. Gray, Trans. Am. Math. Soc. 41 (1969) 465.
[17] See for example: C.H. Tze, Parallel torsions on Lie groups and spontaneous compactification in Kaluza–Klein theories, Yale Preprint YTP 83-02, to be published in Phys. Lett. 128B (1983).
[18] H. Flanders, Differential forms (Academic Press, New York, 1963) sections 7.3–7.6.
[19] P. Libermann, J. Diff. Geom. 5 (1973) 511, and references therein.
[20] R. Finkelstein, J. Math. Phys. 7 (1966) 1633.
[21] T. Eguchi, P. Gilkey and A.J. Hanson, Phys. Rep. 66 (1980) 213.
[22] H.R. Skyrme, Proc. Roy. Soc. (London) 262 (1961) 237;
see also Namik K. Pak and C.H. Tze, Ann. Phys. (NY) 117 (1979) 169, and references therein;
E. Witten, Current algebra, baryons, and quark confinement, Princeton Univ. preprint 1983.
[23] V. Guillemin and A. Pollack, Differential topology (Prentice Hall, London, 1974).
[24] N. Steenrod, The topology of fibre bundles (Princeton U.P., Princeton, 1951) pp. 108–110.
[25] L.J. Boya and J. Matéos, J. Phys. A13 (1980) L285.
[26] F. Gürsey and C.H. Tze, Ann. Phys. (NY) 128 (1980) 29.

UNIFICATION OF YANG–MILLS THEORY AND SUPERGRAVITY IN TEN DIMENSIONS

G.F. CHAPLINE
Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

and

N.S. MANTON
Institute for Theoretical Physics, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

Received 23 August 1982

We show how to generalize the coupling of $n = 1$ super-Maxwell theory and $n = 1$ supergravity in ten dimensions to the case of a non-abelian gauge group. We find that the supergravity 2-form potential $a_{\mu\nu}$ is coupled to the Yang–Mills gauge potential A_μ via the Chern–Simons 3-form.

It has recently been demonstrated [1,2] that realistic theories of quarks and leptons and their interactions can be derived by dimensional reduction of E_8 gauge theories defined on certain ten-dimensional spaces. Since the fermions are assigned to the adjoint **248** representation of E_8, these theories would be supersymmetric in flat ten-dimensional spacetime. It happens, however, that the six-dimensional space of extra dimensions must be a compact coset space [1]. To maintain the supersymmetry of the ten-dimensional gauge theory, it therefore seems necessary to introduce supergravity, which restores supersymmetry as a local invariance [3].

The problem of how to unify supersymmetric Yang–Mills theory and supergravity in four dimensions was solved some time ago [4], but to our knowledge a solution for ten-dimensional spacetime has not yet appeared. However, the unification of supersymmetric Maxwell theory and simple ($n = 1$) supergravity in ten dimensions has been achieved [5], and our results and notation are based on this work. One of the interesting features of the Maxwell-supergravity theory is that a gauge transformation of the Maxwell potential a_μ must be accompanied by a transformation of the 2-form supergravity potential $a_{\mu\nu}$, in order that the lagrangian be gauge invariant.

In this note we show how to generalize the Maxwell-supergravity couplings to obtain a unification of $n = 1$ supergravity and $n = 1$ supersymmetric Yang–Mills theory in ten dimensions. The lagrangian is invariant both under local supersymmetry transformations and local non-abelian gauge transformations. We again find that a gauge transformation must be accompanied by a transformation of the supergravity potential $a_{\mu\nu}$.

The $n = 1$. $D = 10$ supergravity multiplet is $\{e_\mu^m, \psi_\mu, a_{\mu\nu}, \lambda, \phi\}$; respectively a zehnbein, Rarita–Schwinger field, 2-form potential, Dirac field, and a scalar. The supergravity lagrangian is [5,6]:

$$\begin{aligned} L_{\mathrm{SG}} = & -\tfrac{1}{2}eR(e, \omega(e)) - \tfrac{1}{2}e\bar{\psi}_\mu\Gamma^{\mu\rho\sigma}D_\rho(\omega(e))\,\psi_\sigma - \tfrac{3}{4}e\phi^{-3/2}f_{\rho\mu\nu}f^{\rho\mu\nu} - \tfrac{1}{2}e\bar{\lambda}\Gamma^\mu D_\mu(\omega(e))\,\lambda - \tfrac{9}{16}e(\partial_\mu\phi/\phi)^2 \\ & - \tfrac{3}{8}\sqrt{2}\,e\bar{\psi}_\mu(\partial\!\!\!/\phi/\phi)\,\Gamma^\mu\lambda + \tfrac{1}{16}\sqrt{2}\,e\phi^{-3/4}f_{\alpha\beta\gamma}(\bar{\psi}_\mu\Gamma^{\mu\alpha\beta\gamma\nu}\psi_\nu + 6\bar{\psi}^\alpha\Gamma^\beta\psi^\gamma - \sqrt{2}\,\bar{\psi}_\mu\Gamma^{\alpha\beta\gamma}\Gamma^\mu\lambda) \\ & + \text{four-fermion couplings}\,. \qquad (1) \end{aligned}$$

The first term is the Einstein lagrangian, except that the connection ω involves some torsion. e denotes det e_μ^m. The 2-form potential $a_{\mu\nu}$ appears only through the 3-form field $f_{\rho\mu\nu} \equiv \partial_{[\rho}a_{\mu\nu]}$, so the transformation $a_{\mu\nu} \to$

$a_{\mu\nu} + \partial_{[\mu}\Phi_{\nu]}$ has no physical effect. The antisymmetrization, here and elsewhere, is with weight one. Multi-index Γ matrices denote antisymmetrized products of Dirac matrices; e.g. $\Gamma^{\rho\mu\nu} = \Gamma^{[\rho}\Gamma^{\mu}\Gamma^{\nu]}$. Because of the anti-commutation relations, $\Gamma^{\rho\mu\nu} = \Gamma^{\rho}\Gamma^{\mu}\Gamma^{\nu}$, if ρ, μ, ν are all different.

In the Maxwell-supergravity theory, one defines a new 3-form field

$$F^{(\mathrm{Ab})}_{\rho\mu\nu} \equiv f_{\rho\mu\nu} - 2^{-1/2}\kappa\, a_{[\rho}f_{\mu\nu]}\,, \tag{2}$$

to effect the coupling of $a_{\mu\nu}$ to the gauge potential a_μ. Here $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$ and κ is the ten-dimensional gravitational coupling constant. $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ is gauge invariant if one postulates that a gauge transformation $\delta a_\mu = \partial_\mu\Lambda$ is accompanied by a transformation $\delta a_{\mu\nu} = 2^{-1/2}\kappa\,\Lambda f_{\mu\nu}$. The Maxwell-supergravity lagrangian is obtained from the pure supergravity lagrangian by adding a number of manifestly gauge-invariant terms and making the replacement $f_{\rho\mu\nu} \to F^{(\mathrm{Ab})}_{\rho\mu\nu}$ (see ref. [5]).

Now consider the supersymmetry transformation

$$\delta a_\mu = \tfrac{1}{2}\Omega_\mu\,, \tag{3a}$$

and the part of the supersymmetry transformation of $a_{\mu\nu}$ which involves a_μ,

$$\delta a_{\mu\nu} = 2^{-1/2}\kappa\,\Omega_{[\mu}a_{\nu]}\,. \tag{3b}$$

Here $\Omega_\mu \equiv \phi^{3/8}\bar{\epsilon}\Gamma_\mu\chi$, where χ is the Dirac field of the supersymmetric Maxwell theory and ϵ is the supersymmetry parameter. Under these transformations the change in the field strength $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ is

$$\delta F^{(\mathrm{Ab})}_{\rho\mu\nu} = -2^{-1/2}\kappa\,\Omega_{[\rho}f_{\mu\nu]}\,. \tag{4}$$

Since the r.h.s. of (4) is gauge invariant, we expect that both the non-abelian generalization of $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ and its order κ supersymmetry variation should be gauge invariant too.

One might guess that to generalize the $D = 10$ Maxwell-supergravity theory to the non-abelian case one would simply introduce traces wherever products of fields carrying an internal symmetry index appear, replace derivatives $\partial_\mu\,\cdot$ by covariant derivatives $D_\mu\,\cdot = \partial_\mu\,\cdot + g[A_\mu, \cdot]$, and replace the Maxwell field tensor $f_{\mu\nu}$ by its non-abelian counterpart $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu + g[A_\mu, A_\nu]$. However, this is not quite correct. The appropriate generalization of $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ in the Yang–Mills case is

$$F^{(\mathrm{YM})}_{\rho\mu\nu} \equiv f_{\rho\mu\nu} - 2^{-1/2}\kappa\, X_{\rho\mu\nu}\,, \tag{5}$$

where $X_{\rho\mu\nu}$ is the Chern–Simons 3-form:

$$X_{\rho\mu\nu} \equiv \mathrm{Tr}(A_{[\rho}F_{\mu\nu]} - \tfrac{2}{3}gA_{[\rho}A_\mu A_{\nu]})\,.$$

It has the property that

$$\partial_{[\sigma}X_{\rho\mu\nu]} = \tfrac{1}{2}\mathrm{Tr}(F_{[\sigma\rho}F_{\mu\nu]})\,,$$

where the r.h.s. is the second Chern form. In the abelian case one has

$$\partial_{[\sigma}a_\rho f_{\mu\nu]} = \tfrac{1}{2}f_{[\sigma\rho}f_{\mu\nu]}\,,$$

which shows that $a_{[\rho}f_{\mu\nu]}$ is the abelian Chern–Simons form. It is therefore plausible that $X_{\rho\mu\nu}$ should be the non-abelian generalization of $a_{[\rho}f_{\mu\nu]}$.

Under an infinitesimal gauge transformation

$$\delta A_\mu = D_\mu\Lambda\,, \tag{6}$$

the Chern–Simons form $X_{\rho\mu\nu}$ changes. However if we postulate an accompanying transformation

$$\delta a_{\mu\nu} = \sqrt{2}\,\kappa\,\mathrm{Tr}(\Lambda\partial_{[\mu}A_{\nu]})\,, \tag{7}$$

then $\delta F^{(\mathrm{YM})}_{\rho\mu\nu} = 0$. We show below that $F^{(\mathrm{YM})}_{\rho\mu\nu}$ is also invariant under finite gauge transformations, which is non-trivial in the non-abelian case. The supersymmetry transformations analogous to (3a) and (3b) are

$$\delta A_\mu = \tfrac{1}{2}\Omega_\mu\,, \qquad \delta a_{\mu\nu} = 2^{-1/2}\kappa\, \mathrm{Tr}(\Omega_{[\mu}A_{\nu]})\,, \tag{8}$$

where Ω_μ is the same as before, except that χ is now a Dirac field in the adjoint representation of the gauge group. Under these transformations

$$\delta F^{(\mathrm{YM})}_{\rho\mu\nu} = -2^{-1/2}\kappa\, \mathrm{Tr}(\Omega_{[\rho}F_{\mu\nu]})\,. \tag{9}$$

The r.h.s. of (9) is gauge invariant and has the same form as (4) with $F_{\mu\nu}$ replacing $f_{\mu\nu}$.

These results suggest that the lagrangian unifying Yang–Mills theory and supergravity for $D = 10$ can be obtained from the Maxwell-supergravity theory by replacing $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ by $F^{(\mathrm{YM})}_{\rho\mu\nu}$ in addition to making the usual changes from an abelian to a non-abelian gauge theory mentioned before. The supersymmetry transformation laws in the Yang–Mills case should differ from those in the abelian case only to the extent that $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ is replaced by $F^{(\mathrm{YM})}_{\rho\mu\nu}$ and $f_{\mu\nu}$ by $F_{\mu\nu}$. The proposed lagrangian and transformations are

$$\begin{aligned} L = {} & L_{\mathrm{SG}}\,(\text{with } F_{\rho\mu\nu} \text{ replaced by } F^{(\mathrm{YM})}_{\rho\mu\nu}) - \tfrac{1}{4}e\phi^{-3/4}\,\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) - \tfrac{1}{2}e\,\mathrm{Tr}(\bar\chi \not{D}(\hat\omega)\,\chi) \\ & - \tfrac{1}{8}\kappa e\phi^{-3/8}\,\mathrm{Tr}[\bar\chi\Gamma^\mu\Gamma^{\rho\sigma}(F_{\rho\sigma} + \hat F_{\rho\sigma})](\psi_\mu + \tfrac{1}{12}\sqrt{2}\,\Gamma_\mu\lambda) + \tfrac{1}{16}\sqrt{2}\,\kappa e\phi^{-3/4}\,\mathrm{Tr}(\bar\chi\Gamma^{\alpha\beta\gamma}\chi)\,\hat F^{(\mathrm{YM})}_{\alpha\beta\gamma} \\ & - \frac{1}{16\times 96}\sqrt{2}\,\kappa^2 e\,\mathrm{Tr}(\bar\chi\Gamma_{\alpha\beta\gamma}\chi)\,\bar\psi_\mu(4\Gamma^{\alpha\beta\gamma}\Gamma^\mu + 3\Gamma^\mu\Gamma^{\alpha\beta\gamma})\,\lambda - \tfrac{1}{512}\kappa^2 e\,\mathrm{Tr}(\bar\chi\Gamma_{\alpha\beta\gamma}\chi)\,\bar\lambda\Gamma^{\alpha\beta\gamma}\lambda\,; \end{aligned} \tag{10}$$

$$\begin{aligned} & \delta e^m_\mu = \tfrac{1}{2}\bar\epsilon\Gamma^m\psi_\mu\,, \qquad \delta\phi = -\tfrac{1}{3}\sqrt{2}\,\epsilon\lambda\phi\,, \\ & \delta a_{\mu\nu} = \tfrac{1}{4}\sqrt{2}\,\phi^{3/4}(\bar\epsilon\Gamma_\mu\psi_\nu - \bar\epsilon\Gamma_\nu\psi_\mu - \tfrac{1}{2}\sqrt{2}\,\bar\epsilon\Gamma_{\mu\nu}\lambda) + \tfrac{1}{2}\sqrt{2}\,\kappa\phi^{3/8}\bar\epsilon\Gamma_{[\mu}\mathrm{Tr}(\chi A_{\nu]})\,, \\ & \delta\lambda = -\tfrac{3}{8}\sqrt{2}\,(\not{\partial}\phi/\phi)\,\epsilon + \tfrac{1}{8}\phi^{-3/4}\Gamma^{\alpha\beta\gamma}\epsilon\hat F^{(\mathrm{YM})}_{\alpha\beta\gamma} + \frac{1}{12\times 36}\sqrt{2}\,\kappa\,\mathrm{Tr}(\bar\chi\Gamma^{\alpha\beta\gamma}\chi)\,\Gamma_{\alpha\beta\gamma}\epsilon\,, \\ & \delta\psi_\mu = D_\mu(\hat\omega(e,\psi))\,\epsilon + \tfrac{1}{32}\sqrt{2}\,\phi^{-3/4}(\Gamma_\mu^{\alpha\beta\gamma} - 9\delta^\alpha_\mu\Gamma^{\beta\gamma})\,\epsilon\hat F^{(\mathrm{YM})}_{\alpha\beta\gamma} - \frac{1}{16\times 32}(\Gamma_\mu^{\alpha\beta\gamma} - 5\delta^\alpha_\mu\Gamma^{\beta\gamma})\,\epsilon\bar\lambda\Gamma_{\alpha\beta\gamma}\lambda \\ & \qquad + \tfrac{1}{96}\sqrt{2}[(\bar\psi_\mu\Gamma_{mn}\lambda)\,\Gamma^{mn}\epsilon + (\bar\lambda\Gamma_{mn}\epsilon)\,\Gamma^{mn}\psi_\mu + 2(\bar\psi_\mu\lambda)\,\epsilon - 2(\bar\lambda\epsilon)\,\psi_\mu + 4(\bar\psi_\mu\Gamma_m\epsilon)\,\Gamma^m\lambda] \\ & \qquad - \tfrac{1}{256}\kappa\,\mathrm{Tr}(\bar\chi\Gamma^{\alpha\beta\gamma}\chi)(\Gamma_{\mu\alpha\beta\gamma} - 5g_{\mu\alpha}\Gamma_{\beta\gamma})\,\epsilon\,, \\ & \delta A_\mu = \tfrac{1}{2}\phi^{3/8}\bar\epsilon\Gamma_\mu\chi\,, \\ & \delta\chi = -\tfrac{1}{4}\phi^{-3/8}\Gamma^{mn}\hat F_{mn}\epsilon + \tfrac{1}{64}\sqrt{2}\,\kappa\,[3(\bar\lambda\chi)\,\epsilon - \tfrac{3}{2}(\bar\lambda\Gamma^{\alpha\beta}\chi)\,\Gamma_{\alpha\beta}\epsilon - \tfrac{1}{24}(\bar\lambda\Gamma^{\alpha\beta\gamma\delta}\chi)\,\Gamma_{\alpha\beta\gamma\delta}\epsilon]\,. \end{aligned} \tag{11}$$

The caret symbol over a field denotes the supercovariant generalization of that field. For the exact formula, see ref. [5].

Our proof that the lagrangian (10) is supersymmetric depends on showing that all terms occurring in the variation of the lagrangian under a supersymmetry transformation have essentially the same gauge-covariant form in the non-abelian and abelian cases. Since in the abelian theory the net variation vanishes [5], the same will be true in the non-abelian theory. Because the lagrangian and supersymmetry transformations, when expressed in terms of $F^{(\mathrm{Ab})}_{\rho\mu\nu}$ and $F^{(\mathrm{YM})}_{\rho\mu\nu}$, respectively, have the same form in both cases, it is clear that at the algebraic level the variations of the lagrangian do too. However, this argument may not be sufficient since many terms in the variation of L involve partial derivatives of the field variations, and it is not obvious that these occur in the required gauge-covariant combinations.

Instead, we use the general result that if L depends on a number of fields ϕ_i then

$$\delta L = (\delta L/\delta\phi_i)\,\delta\phi_i\,, \tag{12}$$

where $\delta L/\delta\phi_i = 0$ would be a classical field equation. A number of integrations by parts are necessary to obtain (12) and thereby eliminate all derivatives of $\delta\phi_i$. With the exception of $\delta\psi_\mu$ all the field variations are algebraic in ϵ, so this formalism is convenient for comparing the non-abelian and abelian cases.

It is clear from eqs. (10) and (11) that almost all terms in the field equations and field variations will be manifestly covariant and of the same form in the abelian and non-abelian cases. The possible exceptions are those terms arising from variations of $F^{(\mathrm{YM})}_{\rho\mu\nu}$ (or $F^{(\mathrm{Ab})}_{\rho\mu\nu}$) with respect to A_μ (or a_μ) and $a_{\mu\nu}$. To see what happens for these terms, consider the simplified lagrangian

$$\widetilde{L} = -\tfrac{1}{4}\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) - \tfrac{3}{4}F^{(\mathrm{YM})}_{\rho\mu\nu}F^{(\mathrm{YM})\,\rho\mu\nu}\,, \tag{13}$$

and the partial variations δA_μ and $\delta a_{\mu\nu}$ given by (8). In this case, eq. (12) becomes

$$\delta\widetilde{L} = \tfrac{1}{2}\mathrm{Tr}[(D_\mu F^{\mu\rho} + 3\cdot 2^{-1/2}\kappa\, F^{(\mathrm{YM})\,\rho\mu\nu}F_{\mu\nu})\,\Omega_\rho]\,, \tag{14}$$

which is covariant; the non-covariant terms proportional to $(\partial_\rho F^{(\mathrm{YM})\,\rho\mu\nu})\,\Omega_\mu A_\nu$, which occur in the separate variations with respect to A_μ and $a_{\mu\nu}$, cancel. Eq. (14) can also be derived by using eq. (9) to compute the variation of the second term in (13), and then non-covariant terms never appear.

The abelian case is quite analogous. If

$$\widetilde{L} = -\tfrac{1}{4}f_{\mu\nu}f^{\mu\nu} - \tfrac{3}{4}F^{(\mathrm{Ab})}_{\rho\mu\nu}F^{(\mathrm{Ab})\,\rho\mu\nu}\,,$$

and δa_μ and $\delta a_{\mu\nu}$ are given by (3a) and (3b), then

$$\delta\widetilde{L} = \tfrac{1}{2}(\partial_\mu f^{\mu\rho} + 3\cdot 2^{-1/2}\kappa\, F^{(\mathrm{Ab})\,\rho\mu\nu}f_{\mu\nu})\,\Omega_\rho\,,$$

which is of the same form as (14). This result extends easily to the full lagrangian and supersymmetry transformations and completes our argument that the supersymmetry of the Maxwell-supergravity theory implies the supersymmetry of the Yang–Mills generalization, eq. (10).

One remaining question about the Yang–Mills-supergravity theory is whether the lagrangian is invariant under finite gauge transformations. The answer is yes, as can be seen as follows. Under a finite gauge transformation, A_μ changes to

$$A^h_\mu = h^{-1}A_\mu h + g^{-1}h^{-1}\partial_\mu h\,,$$

and the Chern–Simons form $X_{\rho\mu\nu}$ changes to

$$X^h_{\rho\mu\nu} = X_{\rho\mu\nu} - \frac{2}{g}\partial_{[\rho}\mathrm{Tr}(\partial_\mu h h^{-1}A_{\nu]}) - \frac{2}{3g^2}\mathrm{Tr}[(\partial_{[\rho}h)\,h^{-1}(\partial_\mu h)\,h^{-1}(\partial_{\nu]}h)\,h^{-1}]\,.$$

Since both $\partial_{[\sigma}X^h_{\rho\mu\nu]}$ and $\partial_{[\sigma}X_{\rho\mu\nu]}$ are equal to the gauge-invariant quantity $\tfrac{1}{2}\mathrm{Tr}(F_{[\sigma\rho}F_{\mu\nu]})$, $X^h_{\rho\mu\nu} - X_{\rho\mu\nu}$ is a closed 3-form. If the ten-dimensional spacetime has the topology of $\mathbb{R}^{10}$, then $X^h_{\rho\mu\nu} - X_{\rho\mu\nu}$ is also exact; i.e. there exists a 2-form $J_{\mu\nu}$ such that

$$X^h_{\rho\mu\nu} = X_{\rho\mu\nu} + \partial_{[\rho}J_{\mu\nu]}\,. \tag{15}$$

If we postulate that

$$a^h_{\mu\nu} = a_{\mu\nu} + 2^{-1/2}\kappa\, J_{\mu\nu}\,, \tag{16}$$

then the 3-form $F^{(\mathrm{YM})}_{\rho\mu\nu}$ and hence the complete lagrangian (10) is gauge invariant.

It is customary to assume that all fields in a gauge theory decay rapidly at spatial infinity, in which case $\mathbb{R}^{10}$ is effectively compactified to $S^9 \times \mathbb{R}^1$, and gauge transformations $h(x)$ tend rapidly to 1 at infinity. With this topology $X^h_{\rho\mu\nu} - X_{\rho\mu\nu}$ is still exact, since the relevant cohomology group $H_3(S^9)$ is trivial. The interpretation of this result is that when $h \to 1$ at infinity, $J_{\mu\nu}$ can be chosen to vanish sufficiently rapidly at infinity that its surface integrals there are zero.

Suppose now that $h(x)$ is a gauge transformation homotopic to the identity. Then there exists a continuous family of gauge transformations $\widetilde{h}(x,\tau)$, $0 \leqslant \tau \leqslant 1$, such that $\widetilde{h}(x,0) = 1$ and $\widetilde{h}(x,1) = h(x)$. In this case we can give an explicit formula for $J_{\mu\nu}$:

$$J_{\mu\nu} = -\frac{2}{g}\,\mathrm{Tr}(\partial_{[\mu} h h^{-1} A_{\nu]}) + \frac{2}{g^2}\int_0^1 \mathrm{Tr}\left[\partial_{[\mu}\widetilde{h}\widetilde{h}^{-1}\,\frac{\mathrm{d}}{\mathrm{d}\tau}(\partial_{\nu]}\,\widetilde{h}\widetilde{h}^{-1})\right]\mathrm{d}\tau\,, \tag{17}$$

This formula was derived from a more general formula of Chern's [7]. If $h(x) = 1 + g\Lambda(x)$, with Λ infinitesimal, then the natural choice $\widetilde{h}(x, \tau) = 1 + g\tau\Lambda(x)$ gives $J_{\mu\nu} = 2\mathrm{Tr}(A_{[\mu}\,\partial_{\nu]}\,\Lambda)$. Substituting this expression into (16) leads to a transformation law for $a_{\mu\nu}$ which differs from (7) only by a physically irrelevant curl. If the gauge transformation is not homotopic to the identity, then formula (17) cannot be used; however $J_{\mu\nu}$ still exists and formulae (15) and (16) give the transformation laws for $X_{\rho\mu\nu}$ and $a_{\mu\nu}$.

To summarize, we have generalized the Maxwell-supergravity theory in ten dimensions and obtained the Yang–Mills supergravity theory. The appearance of the Chern–Simons 3-form is the most novel feature. This form has appeared before in gauge theories, notably in the analysis of the axial anomaly [8], and also as a mass term in three-dimensional Yang–Mills theory [9].

It is conceivable that Chern–Simons forms, or some generalization to include fermionic fields, may play a much larger role in supergravity theories. Recall that part of the eleven-dimensional supergravity lagrangian is $Y \equiv F \wedge F \wedge A$, where A is a 3-form potential and $F = \mathrm{d}A$ [10]. Y is a Chern–Simons form, because in dimension twelve or greater $\mathrm{d}Y = F \wedge F \wedge F$, and integrals of $F \wedge F \wedge F$ over closed 12-manifolds are topological invariants.

It would be most interesting to consider the dimensional reduction of the Yang–Mills-supergravity theory to four-dimensions. If one used the conventional method, where all fields are simply assumed to be independent of the six extra coordinates, one should obtain an $n = 4$ supergravity theory coupled to six $n = 4$ vector multiplets as well as to an $n = 4$ Yang–Mills multiplet. This theory has been predicted to be finite to all orders in perturbation theory by Kallosh [11]. On the other hand, one could use the more sophisticated dimensional reduction procedure [12], where the extra dimensions form a compact coset space. This would extend the models discussed in refs. [1] and [2] to include gravity. The resulting four-dimensional theories would have less than $n = 4$ supersymmetry and might be more realistic phenomenologically.

It is interesting that the present theory has the same zero mass particle content in ten dimensions as the theory of interacting unoriented supersymmetric strings [13]. However, our theory is not the zero slope limit of the string theory and the dimensional reduction just referred to can be carried out without assuming that the scale size of the space of extra dimensions is extremely small.

We would like to thank Peter van Nieuwenhuizen, John Schwarz and Dick Slansky for helpful conversations. One of the authors (N.M.) would also like to acknowledge the support of NSF grant #PHY77-27084. This work was completed at the Aspen Center for Physics and we are grateful for their hospitality.

References

[1] G. Chapline and R. Slansky, Los Alamos preprint LA-UR-82-1076, to be published in Nucl. Phys. B.
[2] D. Olive and P. West, Imperial College preprint ICTP/81/82-25.
[3] P. van Nieuwenhuizen, Phys. Rep. 68 (1981) 189.
[4] S. Ferrara, F. Gliozzi, J. Scherk and P. van Nieuwenhuizen, Nucl. Phys. B117 (1976) 333;
D.Z. Freedman and J.H. Schwarz, Phys. Rev. D15 (1977) 1007.
[5] E. Bergshoeff, M. de Roo, B. de Wit and P. van Nieuwenhuizen, Nucl. Phys. B195 (1982) 97.
[6] A.H. Chamseddine, Nucl. Phys. B185 (1981) 403.
[7] S.-s. Chern, Complex manifolds without potential theory (Springer, Berlin, 1979) appendix 3.
[8] R. Jackiw, Rev. Mod. Phys. 49 (1977) 681.
[9] S. Deser, R. Jackiw and S. Templeton, Ann. Phys. 140 (1982) 372.
[10] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.
[11] R.E. Kallosh, Lebedev Physical Inst. preprint 69 (1982).
[12] P. Forgács and N.S. Manton, Commun. Math. Phys. 72 (1980) 15.
[13] J.H. Schwarz, Superstring theory, Cal. Tech. preprint CALT-68-911 (1982).

TEN INTO FOUR WON'T GO

D.Z. FREEDMAN [1], G.W. GIBBONS [2] and P.C. WEST [3]
Institute for Theoretical Physics, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

Received 11 February 1983

It is shown that nontrivial spontaneous compactification of ten-dimensional $N = 1$ supergravity with or without Yang–Mills matter is not possible unless maximal symmetry (i.e. Lorentz invariance) is violated in the four-dimensional spacetime.

There has been recent interest in obtaining realistic theories of quarks and leptons by dimensional reduction from ten-dimensional supersymmetric Yang–Mills theories [1,2]. These Yang–Mills theories would be supersymmetric in flat ten-dimensional spacetime but in order to break some of the gauge symmetries the ten-dimensional spacetime is chosen to be nontrivial. In particular, $S^2 \times R^8$ was used in ref. [1]. Such nontrivial backgrounds lead to explicit breakings of supersymmetry in the four-dimensional lagrangians. To remedy this it has been suggested that one should consider the Yang–Mills theory coupled to $N = 1$ supergravity in ten dimensions, in which dimensional reduction would occur spontaneously in the solution of the field equations.

The coupled Yang–Mills supergravity theory in ten dimensions has been constructed by trivial dimensional reduction of eleven-dimensional supergravity [3] to $N = 1$ in ten dimensions [4,5] and then coupling, by the Noether method, an abelian [5] and later a non-abelian [6] matter multiplet.

To implement nontrivial dimensional reduction one would seek a purely bosonic solution of the ten-dimensional equations of motion which is both topologically and metrically a product of the form $M_4 \times M_6$ where M_4 is four-dimensional spacetime with a homogeneous and locally Lorentz invariant metric (i.e. M_4 would be Minkowski spacetime, de Sitter space or anti-de Sitter space) and M_6 is some compact six-dimensional riemannian manifold without boundary. In addition, the background Yang–Mills field strengths F_{MN}, and the background three-form F_{MNP} and scalar ϕ which appear as part of the $N = 1$ supergravity multiplet in ten dimensions should not pick out any preferred directions in the four-dimensional spacetime which would otherwise break the maximal spacetime symmetry of the background. This means that F_{MNP}, F_{MN}, and $\partial_M \phi$ vanish if M, N, or $P = 0, 1, 2, 3$.

The purpose of this note is to point out that these requirements are inconsistent unless $F_{MNP} = 0$ and $F_{MN} = 0$ and ϕ is constant. This effectively rules out nontrivial dimensional reduction in both the matter-coupled and pure supergravity theories.

The bosonic part of the Yang–Mills supergravity lagrangian is, in signature convention $-+++++++++$,

$$L = -\tfrac{1}{2}R - \tfrac{3}{4}\phi^{-3/2}(F'_{MNP})^2 - \tfrac{9}{16}(\partial_M\phi/\phi)^2 - \tfrac{1}{4}\phi^{-3/4}\,\mathrm{tr}(F_{MN})^2, \qquad (1)$$

where F'_{MNP} is the modified three-form introduced in ref. [5].

From (1) it is clear that ϕ is a strictly positive field variable since the lagrangian would otherwise become complex. We may therefore define a new scalar field σ by

$$\phi = \exp(\tfrac{2}{3}\sqrt{2}\sigma). \qquad (2)$$

---

[1] Permanent address: Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[2] Permanent address: DAMTP, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK.

[3] Permanent address: Department of Mathematics, King's College, Strand, London WC2, UK.

The field equation for σ is then

$$\partial_M(\sqrt{-g}\, g^{MN}\partial_N\sigma) + \sqrt{-g}[3\sqrt{2}\exp(-\sqrt{2}\sigma)(F'_{MNP})^2 + (1/4\sqrt{2})\exp(-\sigma/\sqrt{2})\mathrm{tr}(F_{MN})^2] = 0. \qquad (3)$$

Since σ is required to be independent of the four spacetime coordinates we can integrate eq. (3) over the compact manifold M_6 to obtain

$$\int_{M_6} \sqrt{-g}\, d^6y [3\sqrt{2}\exp(-\sqrt{2}\sigma)(F'_{MNP})^2 + (1/4\sqrt{2})\exp(-\sigma/\sqrt{2})\mathrm{tr}(F_{MN})^2] = 0. \qquad (4)$$

Because F'_{MNP} and F_{MN} vanish if M, N, P take the value zero, the quantities $(F'_{MNP})^2$ and $\mathrm{tr}(F_{MN})^2$ are positive. It then follows from (4) that

$$F'_{MNP} = 0, \qquad F_{MN} = 0. \qquad (5, 6)$$

Now multiplying (3) by σ and integrating over M_6 yields

$$\int_{M_6} (\sqrt{-g}\, g^{MN}\, \partial_M\sigma\, \partial_N\sigma)\, d^6y = 0,$$

whence σ = constant. The field equation for g_{MN} now gives

$$R_{MN} = 0. \qquad (7)$$

Restricting M, N in (7) to take values from 5 to 10 shows that M_6 is Ricci flat. This implies that unless M_6 has flat factors it can have no continuous isometries because any Killing vector K^M would have to satisfy

$$\Box K^M = 0 \qquad (8)$$

The $\Box$ operator in (8) is that of M_6 and $M = 5, \ldots, 10$. Thus no internal non-abelian symmetries can be generated from the geometry of M_6. The only solutions of (7) known to us are $S^1 \times S^1 \times K3$ where K3 is the K3 surface with its self-dual metric [7] or the flat six-torus $(S^1)^6$.

In a recent paper [8] the pure supergravity theory in ten dimensions [i.e. bosonic lagrangian (2) with $F_{MN} = 0$] was studied in a different conformal gauge obtained by the transformation $g_{MN} \to \phi^{1/4} g_{MN}$ which brings the lagrangian to the form

$$L = -\tfrac{1}{2}\phi R - \tfrac{3}{4}\phi^{-1}(F_{MNP})^2. \qquad (9)$$

A solution of the resulting field equations of the form $S^3 \times$ (seven-dimensional anti-de Sitter space) was found with F_{MNP} proportional to the volume form on S^3 and ϕ equal to one of the lowest nontrivial S^3 hyperspherical harmonics. A straightforward modification of the argument given above rules out such a solution with $\phi > 0$. Indeed, one easily sees that the hyperspherical harmonic takes both positive and negative values on S^3. We believe that it would not be correct to interpret this solution as a nontrivial dimensional reduction of the ten-dimensional theory because expanding the theory about this background, substituting in (9), and integrating over the internal dimensions gives a divergent action for the small fluctuations of $F_{\mu\nu\rho}$ because of the ϕ^{-1} factor in front of $(F_{MNP})^2$ in (9). Further, the metric in the original conformal gauge of (2) would be complex.

The import of our work is that solutions of the ten-dimensional equations with spontaneous compactification and a maximally symmetric spacetime are ruled out. However, one can relax the condition of maximal symmetry and it is possible to find solutions whose spacetime factor is that of euclidean 2 space times two-dimensional anti-de Sitter space and for which the three-form F_{MNP} has nonvanishing components in spacetime directions. Happily, the situation in eleven-dimensional supergravity [3] is different because the four-index field strength may contain a piece proportional to the volume form of spacetime without violating maximal spacetime symmetry [9]. In addition, there may be other possibilities for nontrivial dimensional reduction, starting from $N = 2$ supergravity in ten dimensions.

This research was supported in part by the National Science Foundation, grant number PHY77-27084.

[1] D. Olive and P. West, Imperial College preprint ICTP/81/82-25.

[2] G. Chapline and R. Slansky, Nucl. Phys. B209 (1982) 461.

[3] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.

[4] A.H. Chamseddine, Nucl. Phys. B185 (1981) 403.

[5] E. Bergshoeff, M. De Roo, B. De Wit and P. Van Nieuwenhuizen, Nucl. Phys. B195 (1982) 97.

[6] G.F. Chapline and N.S. Manton, Phys. Lett. 120B (1983) 105.

[7] S.-T. Yao, Proc. Nat. Acad. Sci. USA 74 (1977) 1798.

[8] M.J. Duff, P.K. Townsend and P. Van Nieuwenhuizen, Phys. Lett. 122B (1983) 232.

[9] P.G.O. Freund and M.A. Rubin, Phys. Lett. 94B (1980) 233.

A NEW SO(7) INVARIANT SOLUTION OF $d = 11$ SUPERGRAVITY

B. de WIT[1]
NIKHEF-H, Amsterdam, The Netherlands

and

H. NICOLAI
CERN, Geneva, Switzerland

Received 26 July 1984

We find a new SO(7) invariant solution of $d = 11$ supergravity by exploiting the relationship of this theory with gauged $N = 8$ supergravity in $d = 4$ dimensions.

Spontaneous compactification of extra dimensions is a possible framework for unifying the fundamental interactions [1]. According to this idea one starts from a higher-dimensional field theory, preferably pure gravity or supergravity, which has a (ground-state) solution with a maximally symmetric $d = 4$ dimensional subspace associated with ordinary space–time and extra dimensions that parametrize a compact manifold that is small enough to prevent their immediate experimental discovery. The resulting theory can usually be rewritten as an effective $d = 4$ field theory of massless fields coupled to infinite towers of massive fields. The $d = 4$ fields arise as coefficient functions in some harmonic expansion about the ground-state solution. These functions depend only on the $d = 4$ space–time coordinates x^μ, whereas the higher-dimensional fields depend also on the extra coordinates y^m. Often one retains only the massless modes in the harmonic expansion. This truncated "low-energy" theory may then have several solutions, which are also solutions of the original higher-dimensional theory. On the other hand most solutions of the latter will have no interpretation in terms of the truncated theory because also some of the fields that are associated with the massive states have acquired nonzero vacuum expectation values.

The above general considerations are relevant when comparing $d = 11$ supergravity [2] to gauged $N = 8$ supergravity in $d = 4$ dimensions [3]. The seven extra dimensions can be compactified to the sphere S^7, and this solution is fully supersymmetric and invariant under the SO(8) isometry group of S^7 and the SO(3, 2) isometry group of the $d = 4$ anti-de Sitter space [4]. Gauged $N = 8$ supergravity has a solution with the same features, which indicates that the S^7 compactification corresponds to gauged $N = 8$ supergravity coupled to massive supermultiplets. This relationship has been confirmed by calculations of the full spectrum of small fluctuations about the S^7 solution, which indeed comprises one massless supermultiplet [4,5], and an infinite tower [6] of massive $N = 8$ anti-de Sitter supermultiplets [7]. Consequently there must be a truncation of $d = 11$ supergravity to pure $N = 8$ supergravity in which all the massive supermultiplets are put to zero. In such a truncation the y-dependence of the $d = 11$ fields is restricted leaving only the $N = 8$ supergravity fields as x-dependent coefficient functions. Knowledge of this truncation could be helpful in elucidating the symmetry structure of gauged $N = 8$ supergravity, and it could lead to new solutions of $d = 11$ supergravity. It is the latter aspect that we intend to explore in this letter. So far two solutions of $d=11$ supergravity have been identified with extrema of the potential of gauged $N=8$ supergravity, namely the compactification of the round [4] and the parallelized [8] S^7. These solutions are invariant under SO(8) and SO(7), respectively. However, $N=8$ supergravity has two solutions with SO(7) and one solution with G_2 symmetry [9], whereas a first search [10] for all $d=11$ solutions with at least G_2 invariance has not led to any other solutions beyond the round and parallelized S^7.

---

[1] Address after 1 September 1984: Institute for Theoretical Physics, University of Utrecht, Utrecht, The Netherlands.

As we have explained elsewhere [11] the truncation of $d=11$ supergravity to gauged $N=8$ supergravity is rather complicated. However, we have been able to determine some systematic features of the embedding of the $N=8$ supergravity fields into the $d=11$ theory. The results have been verified for an SO(7) invariant one-parameter class of field configurations of $N=8$ supergravity in which the pseudoscalar fields acquire an SO(7) invariant constant value. The corresponding field configuration in $d=11$ supergravity has been identified and contains both the round and the parallelized S^7 solutions. The $d=11$ supersymmetry transformation rules at these solutions have been compared to those in $d=4$ dimensions taken at the two solutions of $N=8$ supergravity in the pseudoscalar background, and a complete quantitative agreement was found. However, in $N=8$ supergravity also the scalar fields can acquire SO(7) invariant constant values (the SO(7) groups left invariant by scalars and pseudoscalars correspond to two inequivalent subgroups of SO(8)), and this one-parameter class contains the second SO(7) invariant solution. Consequently, if one can identify the corresponding $d=11$ field configurations one must be able to find a new SO(7) invariant solution of $d=11$ supergravity.

The purpose of this letter is to present this new solution of $d=11$ supergravity. As discussed above we use our knowledge of the relation between the $d=4$ and $d=11$ supergravity fields. For the elfbein field we proceed from the following ansatz

$$E_M{}^A(x,y)=\begin{pmatrix}\Delta^{-1/2}(y)e'^{\alpha}_{\mu}(x) & 0\\ 0 & e_m{}^a(y)\end{pmatrix}, \qquad (1)$$

where $e'^{\alpha}_{\mu}(x)$ is the vierbein associated with the maximally symmetric $d=4$ space–time. We use the notation of ref. [11]. Note that this ansatz is more general than the ones previously considered in the context of Freund–Rubin solutions [12] because of the y-dependent factor $\Delta^{-1/2}(y)$ in (1). From the analysis of the supersymmetry transformation laws it follows that this factor *must* be included for all field configurations corresponding to $N=8$ supergravity which involve deviations of the siebenbein $e_m{}^a$ from the round S^7 background. Those are then parametrized as follows

$$e_m{}^a(y)=\mathring{e}_m{}^b(y)S_{ba}(y), \quad \Delta(y)\equiv\det S_{ab}(y), \qquad (2)$$

where $\mathring{e}_m{}^a(y)$ denotes the S^7 siebenbein with given curvature characterized by the mass parameter m_7. For the four-index field strength we take the ansatz [12]

$$F_{\mu\nu\rho\sigma}=\mathrm{i}f\eta_{\mu\nu\rho\sigma}, \qquad (3)$$

where $\eta_{\mu\nu\rho\sigma}$ is the fully antisymmetric covariant Levi-Civita tensor. Since we have assumed that the $d=4$ subspace is maximally symmetric there is no x-dependence in (2). By invoking the Bianchi identities on (3) it follows that (3) is y-independent, so that f is just a parameter. In contradistinction if we convert to $d=11$ tangent space indices, the tensor $F^{\alpha\beta\gamma\delta}$ does depend on y through the factor $\Delta^{-1/2}$ in (1); $F^{\alpha\beta\gamma\delta}$ is in fact proportional to $f\Delta^2(y)$.

All other fields are zero, so we must now give an ansatz for the metric deviations (2). This ansatz is motivated by our knowledge of the small fluctuations about S^7 corresponding to $N=8$ supergravity excitations [4,5], and the known form of the scalar vacuum expectation value in an SO(7) invariant background [9,13]. This leads us to define the y-dependent vector

$$\xi_a=\tfrac{1}{16}\mathrm{i}C_{IJKL}\bar{\eta}^I\Gamma_{ab}\eta^J\bar{\eta}^K\Gamma^b\eta^L, \qquad (4)$$

where C^{IJKL} denotes the SO(7) invariant self-dual tensor discussed in ref. [13], which satisfies

$$C^{IJKP}C_{LMNP}=6\delta^{IJK}{}_{LMN}+9\delta^{[I}{}_{[L}C^{JK]}{}_{MN]}. \qquad (5)$$

The $\eta^I(y)$ are the Killing spinors of S^7 that satisfy

$$(\mathring{D}_m + \tfrac{1}{2}\mathrm{i}m_7\mathring{e}_m{}^a\Gamma_a)\eta^I(y) = 0 \quad (I = 1,\ldots,8). \tag{6}$$

The vector ξ_a vanishes at certain points on S^7. Its radius is related to a parameter $\xi(y) = \delta^{ab}\xi_{ab}(y)$, where ‡1

$$\xi_{ab} = \tfrac{1}{16}C_{IJKL}\bar{\eta}^I\Gamma_a\eta^J\bar{\eta}^K\Gamma_b\eta^L. \tag{7}$$

ξ varies between -21 and 3, and since $\xi_a^2 = (3-\xi)\times(21+\xi)$ the vector (4) vanishes whenever $\xi = -21$ or 3. This happens at the north and south pole of S^7. Nevertheless it is convenient to express all our results in terms of the unit vector $\hat{\xi}_a$ and ξ. We give the following useful identities

$$\mathring{D}_a\hat{\xi}_b = m_7[(3-\xi)/(21+\xi)]^{1/2}(\delta_{ab} - \hat{\xi}_a\hat{\xi}_b), \tag{8}$$

$$\mathring{D}_a\xi = 2m_7[(3-\xi)(21+\xi)]^{1/2}\hat{\xi}_a, \tag{9}$$

$$\xi_{ab} = \tfrac{1}{6}(3+\xi)\delta_{ab} - \tfrac{1}{6}(21+\xi)\hat{\xi}_a\hat{\xi}_b. \tag{10}$$

To find the correct ansatz for (2) we follow the strategy of ref. [11] and require supersymmetry transformation rules that are consistent upon truncation to pure $N=8$ supergravity. As was explained there, consistency must be achieved by means of a field-dependent (and thus y-dependent) chiral SU(8) transformation. Using this requirement for the supersymmetry variation of the spin-1 fields, it follows that the expression

$$e_a{}^{IJ}(y) = \Delta^{-1/2}S^{-1}{}_{ab}\bar{\eta}^I U^{\mathrm{T}}\Gamma^b U\eta^J, \tag{11}$$

where U is an SU(8) matrix, must satisfy the Killing condition for arbitrary Killing spinors η^I and η^J:

$$\mathring{D}_a e_b{}^{IJ} + \mathring{D}_b e_a{}^{IJ} = 0. \tag{12}$$

After parametrizing S_{ab} and U in terms of (4) it turns out that (12) determines S_{ab} and U modulo two integration constants. The relevant result is the expression for S_{ab} which reads

$$S_{ab} = g^{-1/3}(1+21\tau)^{1/9}[\delta_{ab} + (H^{-1}-1)\hat{\xi}_a\hat{\xi}_b], \tag{13}$$

where g and H are defined by

$$g(\tau,\xi) = [1 + 63\tau^2 - 2\tau(1+9\tau)\xi]^{1/2}, \tag{14}$$

$$H(\tau,\xi) = (1+21\tau)/g(\tau,\xi). \tag{15}$$

Here τ is an arbitrary parameter. Because (12) is a homogeneous equation the result (13) is only determined modulo an arbitrary proportionality constant. However, solutions of the $d=11$ field equation are determined up to an overall scale factor. Therefore the ansatz (13) may contain the same solution several times but not necessarily with the same scale. Indeed, both for $\tau = 0$ and $\tau = -1/9$ the function g is y-independent, and the metric that follows from (13) corresponds to the round sphere, because (13) is simply proportional to an SO(7) matrix. However, the S^7 curvatures are not the same, because the normalization of (13) at $\tau = 0$ and $\tau = -1/9$ is different. Expanding (13) about the round sphere at $\tau = 0$ reproduces the massless fluctuations found before [4,5], namely

$$S_{ab} = (1 - \tfrac{2}{3}\tau)\delta_{ab} + 6\tau(\xi_{ab} - \tfrac{1}{9}\delta_{ab}\xi) + \mathrm{O}(\tau^2). \tag{16}$$

The next step is to calculate the Riemann curvatures corresponding to (1) and (13). The various components are

$$\begin{aligned} R_{mn}{}^{ab} = {} & m_7{}^2\{-2 + \tfrac{2}{9}[(3-\xi)/(21+\xi)] \\ & \times(H^2-1)(H^2+H+3)(H^2-H+3)\} \\ & \times\mathring{e}_m{}^{[a}\mathring{e}_n{}^{b]} + m_7{}^2\{-\tfrac{4}{3}(H^2+H+3) \\ & - \tfrac{4}{9}[(3-\xi)/(21+\xi)](H^2-1) \\ & \times(H^3 - 8H^2 - 3H - 9)\}(H-1)\hat{\xi}_{[m}\hat{\xi}^{[a}\mathring{e}_{n]}{}^{b]}, \end{aligned} \tag{17}$$

---

‡1 It is possible to invert the relations (4) and (7) and find an expression for C_{IJKL} in terms of y-dependent quantities.

$$\begin{aligned} C^{IJKL} = {} & \tfrac{1}{12}(9+\xi)\bar{\eta}^{[I}\Gamma_a\eta^J\bar{\eta}^K\Gamma_a\eta^{L]} \\ & + \tfrac{1}{12}\mathrm{i}\xi_a\bar{\eta}^{[I}\Gamma^{ab}\eta^J\bar{\eta}^K\Gamma_b\eta^{L]} \\ & - \tfrac{1}{4}[\xi_a\xi_b/(3-\xi)]\bar{\eta}^{[I}\Gamma_a\eta^J\bar{\eta}^K\Gamma_b\eta^{L]}. \end{aligned}$$

At $\xi = -21$ this expression reduces to

$$C_{IJKL} = -\Gamma^a{}_{[IJ}\Gamma^a{}_{KL]}.$$

$$
\begin{aligned}
R_{\mu m}{}^{\alpha a} = {} & e'^{\alpha}_{\mu} m_7{}^2 [(3-\xi)/(21+\xi)]\, g \\
& \times \{ -\tfrac{2}{9} H^2 (H^2-1)(H^2+2) \mathring{e}_m{}^a \\
& + \tfrac{2}{3} H (H^2-1) [\tfrac{1}{3} (H-1)(H^2-5H-3) \\
& + (21+\xi)/(3-\xi)] \hat{\xi}_m \hat{\xi}^a \},
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
R_{\mu\nu}{}^{\alpha\beta} = {} & \{ 2 m_4{}^2 \\
& + \tfrac{8}{9} m_7{}^2 g^2 [(3-\xi)/(21+\xi)] H^2 (H^2-1)^2 \} \\
& \times e'^{[\alpha}_{\mu} e'^{\beta]}_{\nu},
\end{aligned}
\tag{19}
$$

where m, n and a, b refer to $d=7$, and μ, ν and α, β to $d=4$ world and tangent space indices, respectively. In (17) and (18) we have used that the vierbein e'^{α}_{μ} parametrizes a maximally symmetric $d=4$ space with curvature proportional to $m_4{}^2$, whereas the background siebenbein $e_m{}^a$ corresponds to the round S^7. Putting $H=1$ in (17) and (18) immediately exhibits the definitions of $m_4{}^2$ and $m_7{}^2$.

After contraction of (17)–(19) with the appropriate elfbein components one finds the $d=11$ Ricci tensor. It is then straightforward to verify (1) and (13) solve the field equations provided that the following relations hold

$$99\tau^2 + 18\tau - 1 = 0, \tag{20}$$

$$f^2 = \tfrac{2}{5} m_7{}^2 (1+21\tau)^{20/3}, \tag{21}$$

$$m_4{}^2 = \tfrac{4}{3} m_7{}^2 (1+21\tau)^2. \tag{22}$$

There are two solutions of (20)–(22), but these differ only by an overall scale factor.

It remains to be shown that this solution is indeed SO(7) invariant. For this purpose we define

$$
\begin{aligned}
K^{mIJ} = {} & \mathring{e}^{ma} [(21+\xi)(\delta_{ab} - \hat{\xi}_a \hat{\xi}_b) \mathrm{i} \bar{\eta}^I \Gamma^b \eta^J \\
& + \xi^b \bar{\eta}^I \Gamma_{ab} \eta^J],
\end{aligned}
\tag{23}
$$

which differs from $e_m{}^{IJ}$ by a term proportional to $\bar{\eta}^I \Gamma_a \eta^J$ which separately satisfies the Killing condition (12). Therefore K^{mIJ} is a Killing vector with respect to the round S^7 background. One may also verify that it is also a Killing vector with respect to the full metric that follows from (13), i.e.

$$\mathrm{D}_m K_n{}^{IJ} + \mathrm{D}_n K_m{}^{IJ} = 0, \tag{24}$$

where the covariant derivative is now computed from the full metric, and the index on K^{mIJ} has been lowered with the full metric. To prove that (23) generates the SO(7) subalgebra of SO(8) we observe that the Killing property with respect to the round S^7 background implies

$$K^{mIJ} = P^{IJ}{}_{KL} \mathrm{i} \bar{\eta}^K \Gamma^a \eta^L \mathring{e}^m{}_a, \tag{25}$$

with constant $P^{IJ}{}_{KL}$. To determine this matrix it is most convenient to choose the point y_0 on S^7 where $\eta^I_\alpha(y_0) = \delta^I_\alpha$. After a little algebra one finds

$$P^{IJ}{}_{KL} \propto \delta^{IJ}{}_{KL} - \tfrac{1}{6} C^{IJ}{}_{KL}. \tag{26}$$

It has been demonstrated in ref. [13] that this is a projector onto an SO(7) subalgebra of SO(8), which proves the assertion.

The crucial element in the construction of the above solution was the relation between $d=11$ and $d=4$ supergravity fields as it has been outlined in ref. [11]. This solution was not found in ref. [10] because the authors restricted themselves to configurations (1) with $\Delta(y)=1$; this was done primarily for practical reasons because a general analysis for unrestricted functions $\Delta(y)$ is extremely complicated (we thank N.P. Warner for a discussion on this point). On the basis of the work of ref. [11] the relation between $\Delta(y)$ and the siebenbein $e_m{}^a(y)$ is known, and rather specific restrictions can be derived for possible ansätze. On the other hand, the fact that the above construction was successful confirms once more the correctness of the approach followed in ref. [11]. In particular this concerns the important role played by chiral SU(8) transformations in defining a consistent truncation to pure $N=8$ supergravity. As a further check on this we have also verified the consistency of the gravitino transformation rule at the above solution. Employing again the SU(8) redefinition one finds consistent transformations provided that f has the same value as prescribed by the field equations (cf. (21)). The gravitino transformation rule can also be compared directly to the transformation at the SO(7) invariant scalar solution of $N=8$ supergravity. Using the results of ref. [13] this leads to

$$g^2 = 2\cdot 5^{1/4}(1-3\tau)^2 m_7^2, \tag{27}$$

where g is the SO(8) coupling constant and τ satisfies (20). Substituting this value of g^2 into the value of the $d=4$ cosmological constant at the scalar solution, one finds

$$\Lambda = -2\cdot 5^{3/4} g^2 = -20(1-3\tau)^2 m_7^2. \tag{28}$$

Using (20) the above result leads to $\Lambda = -3m_4^2$, with m_4^2 as quoted in (22), which indicates that the $d=11$ solution does indeed correspond to the scalar solution of $N=8$ supergravity. In view of this correspondence it follows that the $d=11$ solution must be unstable [13].

The results of this paper indicate that most of the solutions of gauged $N=8$ supergravity are to be found on the basis of (1) and (2), possibly with extra torsion (i.e. $F_{mnpq} \neq 0$). The most interesting solution is the one where both $e_m{}^a$ and F_{mnpq} take SO(7) invariant values, but under mutually inequivalent SO(7) subgroups of SO(8). In that case one expects to obtain a solution with G_2 invariance, because $N=8$ supergravity has such a solution. Interestingly enough the latter has residual $N=1$ supersymmetry [9]. Therefore its $d=11$ counterpart will be the first example of a solution with residual supersymmetry and $F_{mnpq} \neq 0$. However, from a $d=11$ point of view there is no reason to restrict oneself entirely to solutions that are related to $N=8$ supergravity, and we expect that many more interesting solutions will be found.

Note added. It has been emphasized by van Nieuwenhuizen [14] that the ansatz (1), where $\Delta(y)$ is replaced by an arbitrary function of y, represents the most general configuration which allows for maximal $d=4$ symmetry. He has also examined whether this ansatz may lead to a solution of the field equations with vanishing cosmological constant.

References

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys. K1 (1921) 966;
O. Klein, Z. Phys. 37 (1926) 895;
B.S. DeWitt, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965);
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Rep. Math. Phys. 1 (1970) 29;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
J. Scherk and J.H. Schwarz, Phys. Lett. B57 (1975) 463;
E. Cremmer and J. Scherk, Nucl. Phys. B103 (1976) 393; B108 (1976) 409;
A. Salam and J. Strathdee, Ann. Phys. (NY) 141 (1982) 316.

[2] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.

[3] B. de Wit and H. Nicolai, Phys. Lett. 108B (1981) 285; Nucl. Phys. B208 (1982) 323.

[4] M.J. Duff and C.N. Pope, in: Supersymmetry and supergravity '82, eds. S. Ferrara, J.G. Taylor and P. van Nieuwenhuizen (World Scientific, Singapore).

[5] B. Biran, F. Englert, B. de Wit and H. Nicolai, Phys. Lett. 124B (1983) 45.

[6] B. Biran, A. Casher, F. Englert, M. Rooman and P. Spindel, Phys. Lett. 134B (1984) 179;
A. Casher, F. Englert, M. Rooman and H. Nicolai, preprint CERN-TH3794;
E. Sezgin, Phys. Lett. 138B (1984) 57.

[7] D.Z. Freedman and H. Nicolai, Nucl. Phys. B237 (1984) 342.

[8] F. Englert, Phys. Lett. 119B (1982) 339.

[9] N.P. Warner, Phys. Lett. 128B (1983) 169.

[10] M. Günaydin and N.P. Warner, preprint CALT-68-1077.

[11] B. de Wit and H. Nicolai, Nucl. Phys. B243 (1984) 91.

[12] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233.

[13] B. de Wit and H. Nicolai, Nucl. Phys. B231 (1984) 506.

[14] P. van Nieuwenhuizen, Les Houches Lectures 1983 (North-Holland, Amsterdam), to be published.

CHAPTER

5

Towards a Realistic Kaluza-Klein Theory

SEARCH FOR A REALISTIC KALUZA-KLEIN THEORY*

Edward WITTEN

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544, USA

Received 12 January 1981

An attempt is made to construct a realistic model of particle physics based on eleven-dimensional supergravity with seven dimensions compactified. It is possible to obtain an SU(3) × SU(2) × U(1) gauge group, but the proper fermion quantum numbers are difficult to achieve.

In 1921 Kaluza suggested [1] that gravitation and electromagnetism could be unified in a theory of five-dimensional riemannian geometry. The idea was further developed by Klein [2] and was the subject of considerable interest during the classical period of work on unified field theories [3]. Readable expositions of some of the classical work have been given in text books by Bergmann and by Lichnerowicz [4]; more recent discussions have been given by Rayski and by Thirring [5].

While the Kaluza–Klein approach has always been one of the most intriguing ideas concerning unification of gauge fields with general relativity, it has languished because of the absence of a realistic model with distinctive and testable predictions. Yet the urgency of the unification of gauge fields with general relativity has surely greatly increased with the growing importance of gauge fields in physics. Moreover, the Kaluza–Klein theory has generalizations to non-abelian gauge fields which actually were first proposed [6] well before real applications were known for Yang–Mills fields in physics.

In the last few years this approach has been revived by Scherk and Schwarz and by Cremmer and Scherk, originally in connection with dual models [7]. These authors introduced many new ideas as well as new focus. In contrast to much of the classical literature, they advocated that the extra dimensions should be regarded as true, physical dimensions, on a par with the four observed dimensions. Cremmer and Scherk suggested that the obvious differences between the four observed dimensions and the extra microscopic ones could arise from a spontaneous breakdown of the vacuum symmetry, or, as they called it, from a process of "spontaneous compactification" of the extra dimensions.

These ideas have motivated much recent work. The idea of spontaneous compactification has been developed in more detail by Luciani [8]. An interesting idea by Palla [9] about massless fermions in theories with extra compact dimensions

---

* Research partially supported by NSF grant PHY78-01221.

will figure in some of the discussion below. Manton [10] has discussed some questions that arise in trying to generate Higgs fields as components of the gauge field in extra dimensions. The idea of extra hidden dimensions has stimulated much work in supersymmetry theory, including the successful construction of the $N=8$ supergravity theory by Cremmer, Julia and Scherk and by Cremmer and Julia [11]. This work has been generalized to give models with broken sypersymmetry [12].

In many respects, of course, the modern approaches to this subject tend to differ from the classical point of view. In view of the proliferation of new particles in the last thirty years, one may be more willing today than in the past to postulate the infinite number of new degrees of freedom that must exist if extra dimensions really exist. Much of the classical literature focussed on the need to eliminate a massless spin-zero particle that naturally exists in the original Kaluza–Klein theory; the question seems less urgent today because the obvious answer is that quantum mechanical mass renormalization could easily account for the failure to observe this particle (a mass of 10^{-4} eV would make it undetectable). Some of the early work was motivated by the hope that the fifth dimension could provide the hidden variables that would eliminate indeterminacy from quantum mechanics. Despite the many generalizations and changes in emphasis that have occurred, I will refer generically to theories in which gauge fields are unified with gravitation by means of extra, compact dimensions as Kaluza–Klein theories.

It has often been suggested that spontaneous compactification and supergravity could be usefully combined together. The $N=8$ supergravity theory was constructed by "dimensional reduction" starting from an eleven-dimensional theory. In this context, "dimensional reduction" just means that the fields are taken to be independent of seven of the original eleven coordinates, to which physical reality need not be attributed. However, Cremmer and Julia [11] suggested that one might wish to consider seriously the eleven dimensions and interpret seven of them as compact dimensions in the spirit of Kaluza and Klein. This idea has been raised, on occasion, by various other theorists. In this paper, I will describe an attempt – not completely successful, but not completely unsuccessful either – to construct a realistic theory of Kaluza–Klein type, based on eleven-dimensional supergravity.

As discussed by some of the authors mentioned above, from a modern point of view the Kaluza–Klein unified theory of gravitation and electromagnetism is probably best understood as a theory of spontaneous symmetry breaking in which the group of general coordinate transformations in five dimensions is spontaneously broken to the product of the four-dimensional general coordinate transformation group and a local U(1) gauge group.

Let us review how this arises. One considers standard general relativity in five dimensions with the standard Einstein–Hilbert action

$$A=\int \mathrm{d}^5x\sqrt{g}R. \tag{1}$$

Instead of assuming that the ground state of this system is five-dimensional Minkowski space, which we will denote as M^5, one takes the ground state to be the product $M^4 \times S^1$ of four-dimensional Minkowski space M^4 with the circle S^1. The space $M^4 \times S^1$ is, like M^5, a solution of the five-dimensional Einstein equations. Classically it is difficult to decide which of the spaces M^5 and $M^4 \times S^1$ is a more appropriate choice as the ground state, since they both have zero energy, insofar as energy can be defined in general relativity*. Conventionally, one might assume that the ground state is M^5. In the Kaluza–Klein approach one assumes, instead, that the ground state is $M^4 \times S^1$, and the physical spectrum is determined by studying small oscillations around this ground state. One assumes that the radius of the circle S^1 is microscopically small, perhaps of order of the Planck length, and this accounts for why the existence of this fifth dimension is not noted in everyday experience.

The symmetries of the Kaluza–Klein ground state $M^4 \times S^1$ are the four-dimensional Poincaré symmetries, acting on M^4, and a U(1) group of rotations of the circle S^1. These symmetries would be observed as local or gauge symmetries in the apparent four-dimensional world because the whole theory started with the Einstein action (1) which is generally covariant. In fact, if one considers small oscillations around the "ground state" $M^4 \times S^1$, one finds an infinite number of massive excitations, the masses being of order the inverse of the circumference of S^1. One finds also a finite number of massless modes, which presumably would constitute the low-energy physics. The massless modes turn out to be a spin-two graviton and a spin-one photon, which are gauge particles of the symmetries of $M^4 \times S^1$, and a Brans–Dicke scalar.

The ansatz which exhibits the massless modes is the following. The metric tensor of this theory is a five by five matrix $g_{AB}(x^\mu, \phi)$ which in general may depend on the four coordinates x^μ, $\mu = 1 \cdots 4$, of M^4, and on the angular coordinate ϕ of S^1. The massless modes are those for which g_{AB} is a function of x^μ only. One can then write g_{AB} in block form

$$g_{AB}(x^\mu, \phi) = \left(\begin{array}{c|c} g_{\mu\nu}(x) & A_\mu(x) \\ \hline A_\mu(x) & \sigma(x) \end{array}\right), \tag{2}$$

where $g_{\mu\nu}$ is a four by four matrix (the first four rows and columns of g_{AB}), $A_\mu = g_{\mu 5}$, and $\sigma = g_{55}$. Then $g_{\mu\nu}$ is the ordinary metric tensor of the apparent four-dimensional world, and describes a massless spin-two particle; A_μ is the gauge field of the U(1) symmetry, and σ is the Brans-Dicke scalar.

In the classical work on the Kaluza–Klein theory, it is shown that the five-dimensional Einstein action (1), when expanded in terms of $g_{\mu\nu}$, A_μ, and σ (and the other modes, which decouple from these at low energies) contains a four-dimensional Einstein action $\sqrt{g}R^{(4)}$ for $g_{\mu\nu}$, a Maxwell action $F^2_{\mu\nu}$ for A_μ, and the usual kinetic energy for σ. Also, one can readily check that A_μ transforms as a gauge field $A_\mu \to A_\mu + \partial_\mu \varepsilon$ under coordinate transformations of the special type $(x^i, \phi) \to (x^i, \phi + \varepsilon(x^i))$ if the metric g_{AB} is transformed by the standard rule

---

* The definition of energy in general relativity depends on the boundary conditions, so while both M^5 and $M^4 \times S^1$ have zero energy, a comparison between them is meaningless, like comparing zero apples to zero oranges.

$$g_{AB} \to g_{A'B'} \frac{\partial x'^{A'}}{\partial x^A} \frac{\partial x'^{B'}}{\partial x^B} .$$

The Kaluza–Klein theory thus unifies the metric tensor $g_{\mu\nu}$ and a gauge field A_μ into the unified structure of five-dimensional general relativity. This theory is surely one of the most remarkable ideas ever advanced for unification of electromagnetism and gravitation.

The Kaluza–Klein theory, as noted above, also has a non-abelian generalization, which has been extensively discussed over the years. In this generalization, one starts with general relativity in $4+n$ dimensions, possibly with additional matter fields or with a cosmological constant. Instead of assuming the ground state to be M^{4+n}, Minkowski space of $4+n$ dimensions, one assumes the ground state to be a product space $\mathrm{M}^4 \times \mathrm{B}$, where B is a compact space of dimension n. $\mathrm{M}^4 \times \mathrm{B}$ should be a solution of the classical equations of motion, or possibly, as will be discussed later, a minimum of some effective potential.

As in the previous discussion, symmetries of B will be observed as gauge symmetries in the effective four dimensional world. With a suitable choice of B, one may unify an arbitrary gauge group, abelian or non-abelian, with ordinary general relativity, in a $4+n$ dimensional theory.

The ansatz which generalizes (2) is the following. Let ϕ_i, $i = 1 \cdots n$, be coordinates for the internal space B. Let T^a, $a = 1 \cdots N$, be the generators of the symmetry group G of B. Let the action of the symmetry generator T^a on the ϕ_i be $\phi_i \to \phi_i + K_i^a(\phi)$, where $K_i^a(\phi)$ is the "Killing vector" associated with the symmetry T^a. Then the massless excitations of the candidate "ground state" $\mathrm{M}^4 \times \mathrm{B}$ correspond to an ansatz of the following form:

$$g_{AB}(x^\alpha, \phi^k) = \left(\frac{g_{\mu\nu}(x^\alpha)}{\sum_a A_\mu^a(x^\alpha) K_i^a(\phi^k)} \middle| \frac{\sum_a A_\mu^a(x^\alpha) K_i^a(\phi^k)}{\gamma_{ij}(\phi^k)} \right), \tag{3}$$

where γ_{ij} is the metric tensor of the internal space B. The fields $A_\mu^a(x^\alpha)$ are massless gauge fields of the group G. In this way one may obtain the gauge fields of an arbitrary abelian or non-abelian gauge group as components of the gravitational field in $4+n$ dimensions.

One may verify that the $4+n$ dimensional gravitational action really contains the proper kinetic energy term $\sum_a (F_{\mu\nu}^a)^2$. It is also straightforward to check that under infinitesimal coordinate transformations of the special form $(x^\alpha, \phi_i) \to (x^\alpha, \phi_i + \sum_a \varepsilon^a(x^\alpha) K_i^a(\phi))$, which is an x-dependent symmetry transformation of the internal space B, the field $A_\mu^a(x)$ transforms in the expected fashion, $A_\mu^a(x) \to A_\mu^a(x) + D_\mu \varepsilon^a(x)$. Thus, A_μ^a really has the properties expected of an ordinary four-dimensional gauge field. This gauge field is a remnant of the original coordinate invariance group in $4+n$ dimensions, which has been spontaneously broken down to the symmetries of M×B.

As has been noted before, there is a fairly extensive literature on this construction. The case which has been discussed most widely is the case in which B is itself the manifold of some group H. It should be noted that, if H is a non-abelian group, the symmetry group G of the group manifold is not H but H×H, since the group manifold can be transformed by either left or right multiplication. If one starts with general relativity in $4+n$ dimensions, the ansatz (3) will automatically give massless gauge mesons of the full symmetry group H×H.

What problems arise if we try to construct a realistic theory along these lines? Known particle interactions can be described by the gauge group SU(3)×SU(2)× U(1). So the symmetry group G of the compact space B must at least contain this as a subgroup,

$$\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)\subset\mathrm{G}\,. \tag{4}$$

So B must at least have SU(3)×SU(2)×U(1) as a symmetry group.

To be as economical as possible, we may wish to choose B to be a manifold of minimum dimension with an SU(3)×SU(2)×U(1) symmetry. What is the minimum dimension of a manifold which can have SU(3)×SU(2)×U(1) symmetry?

U(1) is the symmetry group of the circle S^1, which has dimension one. The lowest dimension space with symmetry SU(2) is the ordinary two-dimensional sphere S^2. The space of lowest dimension with symmetry group SU(3) is the complex projective space CP^2, which has real dimension four. (CP^2 is the space of three complex variables (Z^1, Z^2, Z^3), not all zero, with the identification $(Z^1, Z^2, Z^3)\simeq(\lambda Z^1, \lambda Z^2, \lambda Z^3)$ for any non-zero complex number λ. CP^2 can also be defined as the homogeneous space SU(3)/U(2).) Therefore, the space $\mathrm{CP}^2\times\mathrm{S}^2\times\mathrm{S}^1$ has SU(3)× SU(2)×U(1) symmetry, and it has $4+2+1=7$ dimensions.

As we will see below, seven dimensions is in fact the minimum dimensionality of a manifold with SU(3)×SU(2)×U(1) symmetry, although $\mathrm{CP}^2\times\mathrm{S}^2\times\mathrm{S}^1$ is not the only seven-dimensional manifold with this symmetry. If, therefore, we wish to construct a theory in which SU(3)×SU(2)×U(1) gauge fields arise as components of the gravitational field in more than four dimensions, we must have at least seven extra dimensions. With also four non-compact "space-time" dimensions, the total dimensionality of our world must be at least $4+7=11$.

This last number is most remarkable, because eleven dimensions is probably the maximum for supergravity. Eleven-dimensional supergravity has been explicitly constructed, and it is strongly believed that supergravity theories do not exist in dimensions greater than eleven. (The reason for this belief is that, on purely algebraic grounds [13], a supergravity theory in $d>11$ would have to contain massless particles of spin greater than two. But there are excellent reasons, both *S*-matrix theoretic [14] and field theoretic [15], to believe that consistent field theories with gravity coupled to massless particles of spin greater than two do not exist.) It is consequently just barely possible to obtain SU(3)×SU(2)×U(1) gauge fields as part of the gravitational field in a supergravity theory, if we use the unique, maximal, eleven-dimensional supergravity theory.

It is certainly a very intriguing numerical coincidence that eleven dimensions, which is the maximum number for supergravity, is the minimum number in which one can obtain $\mathrm{SU(3)\times SU(2)\times U(1)}$ gauge fields by the Kaluza–Klein procedure. This coincidence suggests that the approach is worth serious consideration.

Let us now discuss in more detail the question of why seven dimensions is the minimum number of dimensions for a space with $\mathrm{SU(3)\times SU(2)\times U(1)}$ symmetry – and the related matter of determining all seven-dimensional manifolds with this symmetry.

The space of lowest dimension with any symmetry group G is always a homogeneous space G/H, where H is a maximal subgroup of G. (The space G/H is defined as the set of all elements g of G, with two elements g and g' regarded as equivalent, $g \simeq g'$, if they differ by right multiplication by an element of H, that is, if $g = g'$ with $h \in \mathrm{H}$.) The dimension of G/H is always equal to the dimension of G minus the dimension of H.

In the case $\mathrm{G = SU(3)\times SU(2)\times U(1)}$, the largest dimension subgroup that is suitable is $\mathrm{SU(2)\times U(1)\times U(1)}$. Any larger subgroup of G would contain as a subgroup one of the three factors SU(3), SU(2), or U(1) of G, and this factor would then not have any non-trivial action on G/H – it would not really be a symmetry group of G/H. Since the dimension of $\mathrm{SU(3)\times SU(2)\times U(1)}$ is $8+3+1=12$ and the dimension of $\mathrm{SU(2)\times U(1)\times U(1)}$ is $3+1+1=5$, the dimension of $\mathrm{(SU(3)\times SU(2)\times U(1))/(SU(2)\times U(1)\times U(1))}$ is $12-5=7$. It is for this reason that a space with $\mathrm{SU(3)\times SU(2)\times U(1)}$ symmetry must have at least seven dimensions. However, there are many ways to embed $\mathrm{SU(2)\times U(1)\times U(1)}$ in $\mathrm{SU(3)\times SU(2)\times U(1)}$, and as a result there are many seven-dimensional manifolds with $\mathrm{SU(3)\times SU(2)\times U(1)}$ symmetry.

To embed $\mathrm{SU(2)\times U(1)\times U(1)}$ in $\mathrm{SU(3)\times SU(2)\times U(1)}$ we first embed SU(2). SU(2) can be embedded in $\mathrm{SU(3)\times SU(2)\times U(1)}$ in a variety of ways. The only embedding that turns out to be relevant is for SU(2) to be embedded in SU(3) as an "isospin" subgroup, so that the fundamental triplet of SU(3) transforms as $2+1$ under SU(2). [Other embeddings of SU(2) lead to spaces G/H on which some of the $\mathrm{SU(3)\times SU(2)\times U(1)}$ symmetries act trivially, as discussed in the previous paragraph.] We still must embed $\mathrm{U(1)\times U(1)}$ in $\mathrm{SU(3)\times SU(2)\times U(1)}$.

$\mathrm{SU(3)\times SU(2)\times U(1)}$ has three commuting U(1) generators which commute with the SU(2) subgroup of SU(3) that we have just chosen. There is a "hypercharge" generator of SU(3), which we may call λ_8, which commutes with the "isospin" subgroup. Also, we have the U(1) factor of $\mathrm{SU(3)\times SU(2)\times U(1)}$, which will be called Y, and we may choose an arbitrary U(1) generator of the SU(2) factor, which will be called T_3.

So $\mathrm{SU(3)\times SU(2)\times U(1)}$ contains an essentially unique subgroup $\mathrm{SU(2)\times U(1)\times U(1)\times U(1)}$, where the three U(1) factors are λ_8, T_3, and Y. We do not want to divide

SU(3) × SU(2) × U(1) by the full SU(2) × U(1) × U(1) × U(1) subgroup because this would yield a space ($CP^2 \times S^2$, to be precise) on which the U(1) of SU(3) × SU(2) × U(1) would act trivially and would not really be a symmetry. So we delete one of the three U(1) factors, and divide only by SU(2) × U(1) × U(1).

The U(1) factor that is deleted may be an arbitrary linear combination $p\lambda_8 + qT_3 + rY$ of λ_8, T_3, and Y where p, q, and r are any three integers which have no common divisor*. So we define H as SU(2) × U(1) × U(1), where the SU(2) is our "isospin" subgroup of SU(3), and the two U(1)'s are the two linear combinations of λ_8, T_3, and Y which are orthogonal to $p\lambda_8 + qT_3 + rY$. The space G/H is then a seven-dimensional space with SU(3) × SU(2) × U(1) symmetry, which we may call M^{pqr}.

In a few cases the M^{pqr} are familiar spaces. M^{001} is our previous example $CP^2 \times S^2 \times S^1$. But in most cases the M^{pqr} are not familiar spaces, and are not products.

In a few cases the M^{pqr} have greater symmetry than SU(3) × SU(2) × U(1). M^{101} is $S^5 \times S^2$, which has the symmetry O(6) × SU(2). M^{011} is $CP^2 \times S^3$, whose full symmetry is SU(3) × SU(2) × SU(2). Except for these two cases, one cannot obtain from seven extra dimensions a symmetry "larger" than SU(3) × SU(2) × U(1). Therefore, the observed gauge group in nature is practically the "largest" group one could obtain from a Kaluza–Klein theory with seven extra dimensions.

Although the M^{pqr} for general values of p, q, and r are not familiar spaces, it is possible to give a rather explicit description of them. Consider first the eight dimensional space $S^5 \times S^3$ [S^n is the n-dimensional sphere, with symmetry group $O(n+1)$]. The symmetry group of $S^5 \times S^3$ is O(6) × O(4). Let us introduce a particular generator of O(6),

$$K = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}, \tag{5}$$

and a particular generator of O(4),

$$L = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{6}$$

Then the subgroup of O(6) that commutes with K is SU(3) × U(1) [the U(1) being generated by K itself] and the subgroup of O(4) that commutes with L is SU(2)×U(1) [the U(1) being generated by L].

---

* And r should be non-zero to avoid obtaining a space on which U(1) is realized as the identity.

For any non-zero p and q, we now define $N = -qK + pL$. Then N generates a U(1) subgroup of O(6)×O(4), consisting of elements of the form exp tN, $0 \leq t \leq 2\pi$. We may now form from $S^5 \times S^3$ a seven-dimensional space $M^{pq} = (S^5 \times S^3)/U(1)$, where two points in $S^5 \times S^3$ are considered to be identical if they are mapped into each other by the action of the U(1) subgroup generated by N.

This space M^{pq} is equal to the $r = 1$ case of what we have previously called M^{pqr}. The M^{pq} are actually the most general simply connected seven-dimensional manifolds with SU(3)×SU(2)×U(1) symmetry. To obtain M^{pqr} for $r \neq 1$ one must factor out from $S^5 \times S^3$ an additional discrete subgroup consisting of elements of the form $\exp(2\pi qK/r)$ $(q = 0, 1, 2, \ldots, r-1)$. We define $M^{pqr} = M^{pq}/Z^r = (S^5 \times S^3)/(U(1) \times Z^r)$.

To verify that the construction of the M^{pqr} just presented is equivalent to the previous definition as (SU(3)×SU(2)×U(1))/(SU(2)×U(1)×U(1)), one uses the fact that SU(3)/SU(2) is S^5, while SU(2) is S^3, so (SU(3)×SU(2)×U(1))/(SU(2)×U(1)×U(1)) is $(S^5 \times S^3 \times U(1))/(U(1) \times U(1))$. Dividing out the two U(1) factors, one arrives at the above definition of M^{pqr} as $(S^5 \times S^3)/(U(1) \times Z^r)$.

The M^{pqr} are not quite the most general seven-dimensional manifold with SU(3)×SU(2)×U(1) symmetry, because for special values of p, q, and r it is possible to supplement SU(2)×U(1)×U(1) with an additional twofold discrete symmetry. One obtains in this way some non-orientable manifolds with one of the M^{pqr} as a double covering space. These spaces are the following. Dividing M^{001} by a discrete symmetry one can get $CP^2 \times P^2 \times S^1$ (P^k is real projective space of dimension k), or $CP^2 \times (S^2 \times S^1)/Z_2$, where Z_2 is a simultaneous inversion of S^2 and S^1. From M^{101} one gets $S^5 \times P^2$ and $(S^5 \times S^2)/Z_2$, where the Z_2 is a simultaneous inversion of S^5 and S^2. Likewise, by dividing M^{10r} by an additional two-fold symmetry one can make $S^5/Z^r \times P^2$ and $(S^5/Z^r \times S^2)/Z_2$. These spaces are non-orientable. This completes the list of seven-dimensional manifolds with SU(3)×SU(2)×U(1) symmetry.

If one is willing to suppose that the ground state of eleven-dimensional supergravity is a product of four-dimensional Minkowski space with one of the M^{pqr}, one can obtain an SU(3)×SU(2)×U(1) gauge group, the gauge fields being components of the gravitational field, according to the ansatz of eq. (3). Of course, to describe nature, it is not sufficient to have the gauge group. It is also necessary to have quarks and leptons of essentially zero mass [very light compared to the energy scale of gravitation; massless in any approximation in which SU(3)×SU(2)×U(1) is not spontaneously broken] which should be in the appropriate representation of the gauge group. And it is necessary to find Higgs bosons whose vacuum expectation value could ultimately trigger SU(2)×U(1) breaking.

How can one obtain massless quarks and leptons in the Kaluza–Klein framework? To understand the basic idea*, suppose that in a $4+n$ dimensional theory we have a massless spin one half fermion. It satisfies the $4+n$ dimensional Dirac equation,

---

* See also a discussion by Palla [9].

$$\not{D}\psi = 0\,, \tag{7}$$

or explicitly

$$\sum_{i=1}^{4+n} \gamma^i D_i \psi = 0\,. \tag{8}$$

This Dirac operator can be written in the form

$$\not{D}^{(4)}\psi + \not{D}^{(\mathrm{int})}\psi = 0, \tag{9}$$

where $\not{D}^{(4)} = \sum_{i=1}^{4} \gamma^i D_i$ is the ordinary four-dimensional Dirac operator, and $\not{D}^{(\mathrm{int})} = \sum_{j=5}^{4+n} \gamma^j D_j$ is the Dirac operator in the internal space of n compact dimensions.

The expression (9) immediately shows that the eigenvalue of $\not{D}^{(\mathrm{int})}$ will be observed in practice as the four-dimensional mass. If $\not{D}^{(\mathrm{int})}\psi = \lambda\psi$, then ψ will be observed by four-dimensional observers who are unaware of the existence of the extra microscopic dimensions as a fermion of mass $|\lambda|$.

The operator $\not{D}^{(\mathrm{int})}$ acts on a compact space, so its spectrum is discrete. Its eigenvalues either are zero or are of order $1/R$, R being the radius of the extra dimensions. Since $1/R$ is, in the Kaluza–Klein approach, presumably of order the Planck mass, the non-zero eigenvalues of $\not{D}^{(\mathrm{int})}$ correspond to extremely massive fermions which would not have been observed. The observed quarks and leptons must correspond to the zero modes of $\not{D}^{(\mathrm{int})}$.

If, in eleven-dimensional supergravity, the ground state is a product of four-dimensional Minkowski space with one of the M^{pqr}, then the zero modes of the Dirac operator in the internal space will, if there are any zero modes at all, automatically form multiplets of $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$, since this is the symmetry of the internal space. It therefore is reasonable to wonder whether for an appropriate choice of p, q, and r, zero modes could exist and form the appropriate representation of the symmetry group, so as to reproduce the observed spectrum of quarks and leptons.

Of course, to reproduce what is observed in nature, we would need quite a few zero modes of the internal space Dirac operator. If the top quark exists, there are in nature at least 45 fermion degrees of freedom of given helicity, counting all colors and flavors of quarks and leptons. We would therefore need at least 45 Dirac zero modes. However, when a Dirac operator has zero modes, the number usually depends on topological invariants. Perhaps by choosing suitable values of p, q, and r we could suitably "twist" the topology and obtain the required 45 zero modes lying in the appropriate representation of $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$.

Actually, if one has in mind eleven-dimensional supergravity, one must modify this program slightly. In eleven-dimensional supergravity, there is no fundamental spin one half field. The only fundamental Fermi field in that theory is the Rarita–Schwinger field $\psi_{\mu\alpha}$, of spin $\frac{3}{2}$ (μ is a vector index, α a spinor index).

Although this field has spin $\frac{3}{2}$ from the point of view of eleven dimensions, the components of ψ_μ with $5 \leq \mu \leq 11$ are spin one half fields from the point of view of ordinary four-dimensional physics. For $\mu \geqslant 5$, μ would be observed as an internal symmetry index, not a space-time index; it carries spin zero. Although the components ψ_μ with $\mu = 1 \cdots 4$ are spin-$\frac{3}{2}$ fields in the four-dimensional sense, the components with $\mu = 5 \cdots 11$ are spin one half fields. So zero-mode solutions of the spin-$\frac{3}{2}$ wave equation in the extra dimensions would be observed as massless spin-$\frac{1}{2}$ fermions in four dimensions. These would be the ordinary light fermions of the spontaneously compactified eleven-dimensional theory.

In one sense, it is an advantage to have to consider the Rarita–Schwinger operator rather than the Dirac operator. The Rarita–Schwinger operator can have zero modes more easily and in more abundance than the Dirac operator, because the Dirac operator has positivity properties which tend to suppress the number of zero modes. For instance, with four extra dimensions, it is known [16] that there is only a single non-flat compact solution of Einstein's equations on which the Dirac operator has zero modes. This is the Kahler manifold K3 (which has no Killing vectors). On this space there are two zero modes of the Dirac operator – but 42 zero modes of the Rarita–Schwinger operator. The large discrepancy is caused, in this case, by a much larger coefficient of the axial anomaly for Rarita–Schwinger fields. This example shows, incidentally, that the rather large number of zero modes that would be required to describe what is observed in physics is not necessarily out of reach.

In the approach considered here, the solution of the problem of flavor – the problem of the existence of several "generations" of fermions with the same quantum numbers – would be that the extra dimensions have a sufficiently complex topology that there are several zero modes with the same $SU(3) \times SU(2) \times U(1)$ quantum numbers. When an operator has several zero modes, they are not necessarily related by any symmetry. For instance, the isospinor Dirac operator in a Yang–Mills instanton of topological number K has K modes; these modes are not related by any symmetry. This is fortunate, because the various generations of fermions have very different masses and are not obviously related by any symmetry.

Unfortunately, there is a basic reason that this idea does not work, at least not in the form described above. The reason for this is related to one of the most basic facts about the observed quarks and leptons: the fermions of given helicity transform in a complex representation of the gauge group, or, to put it differently, right-handed fermions do not transform the same way that the left-handed fermions transform. For instance, left-handed color triplets (quarks) are SU(2) doublets, but right-handed color triplets are SU(2) singlets. This is the reason that quarks and leptons do not have bare masses but receive their mass from the Higgs mechanism – from $SU(2) \times U(1)$ symmetry breaking. This is a very important fact theoretically, because it is the basis for our theoretical understanding of why the quarks and leptons are very light compared to the mass scale of grand unification or the Planck mass. If left- and right-handed fermions transformed the same way under the the gauge group, bare masses would have been possible and could have been arbitrarily large.

In the framework that has been described above, right- and left-handed fermions would inevitably transform the same way under SU(3)×SU(2)×U(1). The reason for this is that low mass fermions are supposed to arise as zero modes of the Rarita–Schwinger operator in the extra dimensions. But the Rarita–Schwinger operator in the seven extra dimensions does not "know" whether a spinor field is left- or right-handed with respect to four-dimensional Lorentz transformations. It treats four-dimensional left- and right-handed fermions in the same way. One therefore could not get the observed SU(3)×SU(2)×U(1) representation. One would inevitably get vector-like rather than V–A weak interactions, with bare masses being possible for all fermions. (Indeed, precisely because bare masses would be possible for all fermions, it is not natural to get any massless fermions at all.)

There is an intriguing mechanism by which, at first sight, it seems that the internal space Rarita–Schwinger equation could treat left and right fermions differently. Eleven-dimensional spinors are constructed with eleven gamma matrices γ_i, $i = 1 \cdots 11$. Let us define an operator $\Gamma_{11} = i\gamma_1 \cdots \gamma_{11}$ which is a sort of eleven-dimensional helicity operator. Let us also define an operator $\Gamma_4 = i\gamma_1\gamma_2\gamma_3\gamma_4$ which measures the ordinary four-dimensional helicity, and an operator $\Gamma_7 = \gamma_5 \cdots \gamma_{11}$ which one might think of as "helicity" in the internal eleven-dimensional space. Then $\Gamma_{11}^2 = \Gamma_4^2 = \Gamma_7^2 = 1$ and $\Gamma_{11} = \Gamma_4\Gamma_7$.

The Rarita–Schwinger field ψ of eleven-dimensional supergravity satisfies a Weyl condition $\psi = \Gamma_{11}\psi$. (This condition must be imposed; otherwise there would be more Fermi than Bose degrees of freedom and supersymmetry would not be possible.) This identity may equivalently be written $\Gamma_4\psi = \Gamma_7\psi$.

The latter equation shows that in eleven-dimensional supergravity the four-dimensional helicity of fermions is correlated with the seven-dimensional "helicity". Components with $\Gamma_4 = +1$ (or -1) have $\Gamma_7 = +1$ (or -1). If the quantum numbers of zero modes of the seven-dimensional Rarita–Schwinger equation depended on Γ_7, as one might intuitively expect, they would also depend on Γ_4.

Unfortunately, the spectrum of the seven-dimensional Rarita–Schwinger operator does not depend on Γ_7. The reason for this is very simple (and depends only on the fact that the number of extra dimensions is odd). In defining how spinors transform under coordinate transformations in riemannian geometry one needs the matrices $\sigma_{ij} = [\gamma_i, \gamma_j]$. One does not (on an orientable manifold) need the γ_i themselves. The transformation $\gamma_i \leftrightarrow -\gamma_i$ does not change the σ_{ij}, so it does not affect the definition of spinors. It does, however, change the sign of $\Gamma_7 = \gamma_1\gamma_2 \cdots \gamma_7$. Consequently, spinors with opposite values of Γ_7 transform the same way under coordinate transformations. Since, in the approach discussed here, SU(3)×SU(2)×U(1) transformations are coordinate transformations, spinors with opposite values of Γ_7 have the same SU(3)×SU(2)×U(1) quantum numbers.

One could try to avoid this conclusion by taking the extra seven dimensions to be a non-orientable manifold. In a non-orientable manifold, the definition of spinors is subtle and involves the γ_i as well as σ_{ij}. However, seven-dimensional non-orientable manifolds with $SU(3)\times SU(2)\times U(1)$ symmetry are not abundant (they have all been listed above), and it is not difficult to show that none of them are suitable.

One might also try to avoid the above stated conclusion by going beyond riemannian geometry to include some variant of torsion. What possibilities this would offer is not very clear; the matter will be discussed at the end of this paper.

Obtaining the right quantum numbers for quarks and leptons is, of course, not the only problem that must be faced in order to obtain a realistic theory, although it may be the most difficult problem. We must also worry about spontaneous breaking of supersymmetry, spontaneous breaking of *CP*, spontaneous breaking of $SU(2)\times U(1)$ gauge symmetry, and obtaining the proper values of the low-energy parameters (coupling constants, masses, and mixing angles); and we must worry about what the true ground state of the theory really is. These questions will now be briefly discussed in turn.

For spontaneous breaking of supersymmetry the prospects are very bright; in fact, supersymmetry almost inevitably is spontaneously broken as part of any scheme in which there are compact dimensions with a non-abelian symmetry.

The reason for this is the following. Unbroken supersymmetry means that under a supersymmetry transformation the vacuum expectation values of the fields do not change. The vacuum expectation values of the Bose fields automatically are invariant under supersymmetry, since their supersymmetric variation would be proportional to the (vanishing) vacuum expectation values of the Fermi fields. The delicate question is whether the vacuum expectation values of the fermi fields change under supersymmetry.

To illustrate the point, let us ignore the possible presence in the theory of Bose fields other than the gravitational field. Then the transformation law for the Rarita–Schwinger field is $\delta\psi_\mu = D_\mu\varepsilon$, ε being the gauge parameter. An unbroken supersymmetry – a symmetry of the vacuum – must have $\delta\psi_\mu = 0$, so unbroken supersymmetry transformations correspond to solutions of $D_\mu\varepsilon = 0$.

On a curved manifold, this equation will almost certainly not have solutions, since $D_\mu\varepsilon = 0$ implies the integrability condition $[D_\mu, D_\nu]\varepsilon = 0$ or $R_{\mu\nu\alpha\beta}[\gamma^\alpha, \gamma^\beta]\varepsilon = 0$, which on most curved manifolds is not satisfied by any non-zero ε. For instance, on none of the M^{pqr} does a solution exist. (The properties of seven-dimensional manifolds admitting solutions of $D_\mu\varepsilon = 0$ have been discussed in the mathematical literature [17], but non-trivial examples do not seem to be known.) So in theories with curved extra dimensions, there will generally not be any unbroken supersymmetries.

The picture does not change greatly when one includes Bose fields other than the gravitational field. We now have $\delta\psi_\mu = \bar{D}_\mu\varepsilon$, where $\bar{D}_\mu = D_\mu$ plus non-minimal terms involving the vacuum expectation values of other Bose fields (and possibly involving the expectation values of fermion bilinears, as discussed below). Unbroken supersymmetries are now solutions of $\bar{D}_\mu\varepsilon = 0$, but solutions will still typically not exist because the integrability condition $[\bar{D}_\mu, \bar{D}_\nu]\varepsilon = 0$ will still not have solutions.

Although solutions will generally not exist, the extra dimensions and the vacuum expectation values of the fields may be just such that one or more solutions of $\bar{D}_\mu \varepsilon = 0$ would exist. Each solution of $\bar{D}_\mu \varepsilon = 0$ in the internal space would correspond to an unbroken supersymmetry charge in four dimensions. If there is precisely one such solution, and so only one unbroken supersymmetry generator, this corresponds to a theory in which $N = 8$ supersymmetry has been spontaneously broken down to $N = 1$ supersymmetry. If there are K solutions, there is an unbroken $N = K$ supersymmetry.

A particularly attractive possibility would be a theory in which the equation $\bar{D}_\mu \varepsilon = 0$ has precisely one solution in the extra dimensions, corresponding to unbroken $N = 1$ supersymmetry. With $N = 1$ supersymmetry it is possible to construct more or less realistic models of observed particle physics. With $N \geqslant 2$ it is not possible to make a realistic model, because the supersymmetry algebra for $N \geqslant 2$ forces left- and right-handed fermions to transform in the same way under the gauge group, in contrast with what is observed. It is attractive to believe that $N = 1$ supersymmetry might survive after compactification of seven dimensions because this would severely constrain the theory, would make many predictions that might be testable in accelerators, and [19] might shed light on SU(2) × U(1) breaking and the gauge hierarchy problem. Of course, we would then have to explain how $N = 1$ supersymmetry is eventually spontaneously broken at low energies.

In addition to supersymmetry breaking, we must also explain P and CP breaking in order to construct a realistic theory. The eleven-dimensional supergravity langrangian is invariant under inversions of space (or time) combined with a change of sign of the antisymmetric tensor gauge field that exists in this theory. After compactification of seven dimensions, the eleven-dimensional symmetry could be manifested as both P (inversion of space) and C (inversion of the compact dimensions). These potential invariances must be spontaneously broken.

A natural mechanism for spontaneous breaking of P, C, and CP involves the antisymmetry tensor gauge field of the eleven-dimensional supergravity theory. The curl $F_{\alpha\beta\gamma\delta}$ of this field may have a vacuum expectation value without breaking Lorentz invariance or $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. In fact, as discussed recently by several authors [20], a vacuum expectation value of F_{1234} is Lorentz invariant. It would violate P and CP but conserve C. The components F_{ijkm} for $i \cdots m \geqslant 5$ may also have expectation values, which would spontaneously break C and CP but conserve P. It is not difficult to see (by considering the little group of a point on M^{pqr}) that on any of the M^{pqr}, the most general $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$ invariant vacuum expectation value of F_{ijkl} depends on two real parameters.

Although the eleven-dimensional theory can have spontaneous breaking of C, P, and CP, the strong interaction angle θ will inevitably vanish at the tree level. The reason for this is that in the eleven-dimensional theory, there is no operator which might be added to the lagrangian which reduces in four dimensions to $\theta \int \mathrm{d}^4 x F_{\mu\nu} \tilde{F}_{\mu\nu}$. There simply does not exist in eleven dimensions any topological invariant that can be written as the integral of a lagrangian density. Of course, the question of how large a vacuum angle might be generated by quantum corrections must wait until we understand how to do calculations in this (presumably) non-renormalizable theory.

It is also necessary, of course, to obtain SU(2) × U(1) symmetry breaking; this presumably means that we must find, at the tree level, a massless Higgs doublet which could later obtain a very tiny negative mass squared.

There are various ways that, in a Kaluza–Klein theory, one might obtain massless charged scalars. In the original Kaluza–Klein theory, with a single compact dimension (a circle) there is a massless scalar (at least at the tree level) because the classical field equations do not determine the radius of the circle. Space-time dependent fluctuations of this radius would be observed as a massless scalar degree of freedom.

If the equations that determine our hypothetical ground state $M^4 \times M^{pqr}$ admit not a unique solution for the metric of M^{pqr} but a whole family of solutions, then oscillations within this family would be observed as massless scalars. Some of these oscillations might involve departures from SU(2) × U(1) symmetry and could be the desired Higgs bosons.

One might also obtain massless scalars as components of the antisymmetric tensor gauge field. In fact, massless scalars can be obtained in this way, but tend to be neutral under the gauge group.

Regardless of where the scalars come from, why would they be massless? The most plausible explanation would be an unbroken supersymmetry relating the massless bosons to massless fermions. This could involve the possibility discussed above that the equation $\bar{D}_\mu \varepsilon = 0$ has a unique non-trivial solution, leaving $N = 1$ supersymmetry unbroken. In this case, of course, we must hope to find a non-perturbative mechanism spontaneously breaking the supersymmetry and giving a small vacuum expectation value to the scalar bosons. (Some relevant issues will be discussed in a future paper [21].)

Without understanding the Higgs bosons and the low-energy symmetry breaking, it is of course not possible to predict the quark and lepton masses and mixing angles. If we understood the dynamics that determines the metric of M^{pqr} (assuming that the ground state really is $M^4 \times M^{pqr}$), we could predict the strong, weak, and electromagnetic coupling constants, since the gauge fields all arise, by the ansatz of eq. (3), as part of the metric tensor in eleven dimensions, and the gauge field kinetic energy is part of the Einstein action. [The most general SU(3) × SU(2) × U(1) invariant metric on M^{pqr} depends on three arbitrary parameters. If we understood the dynamics and could calculate the three parameters, we could predict the SU(3), SU(2), and U(1) coupling constants.] Even though we do not understand this dynamics (see below), it is possible to make a useful comment.

In a theory of this kind, the gauge coupling constants, which are determined by integrating the action over the compact dimensions, would scale as a rather high power of $1/(M_{\mathrm{p}}R)$, where M_{p} is the Planck mass and R is the radius of the extra dimensions. The fact that the observed gauge coupling constants in nature differ from one by only one or two orders of magnitude shows that R cannot be too much greater than $1/M_p$; the extra dimensions really have a radius not too different from 10^{-33} cm.

The eleven-dimensional supergravity theory has no global symmetry that could be interpreted as baryon number, so in this theory nucleons are almost surely unstable. The mass scale in nucleon decay, however, would probably be $1/R$, which is the mass scale of the heavy quanta in this theory. Since, as just noted, $1/R$ cannot be much less than M_p, the nucleon lifetime will probably be very long, perhaps 10^{45} years, which is far too long for nucleon decay to be observable. If the present nucleon decay experiments give a positive result, the approach described in this paper would become significantly less attractive.

It is now time to finally discuss the question of whether one can really sensibly expect $M^4 \times M^{pqr}$ to be the ground state of this theory.

The most attractive possibility would be that $M^4 \times M^{pqr}$ might be a solution of the classical equations of motion, possibly with a suitable vacuum expectation assumed for $F_{\mu\nu\alpha\beta}$. Unfortunately, a straightforward calculation shows that this is not true (regardless of what vacuum expectation value one assumes). If one arbitrarily adds to the lagrangian a cosmological constant (with a sign corresponding to a positive energy density) then $M^4 \times M^{pqr}$ can be a solution. However, local supersymmetry does not permit a cosmological constant in the eleven-dimensional lagrangian.

This problem is not necessarily fatal, since one can always hope that $M^4 \times M^{pqr}$, although not a solution of the classical equations of motion, is the minimum of the appropriate effective potential. In eleven-dimensional supergravity, there is no small dimensionless parameter whose smallness could justify the use of the classical field equations as an approximation. So the fact that $M^4 \times M^{pqr}$ does not satisfy the classical equations, while not encouraging, is not necessarily critical.

In any case, there is absolutely no obvious reason that $M^4 \times M^{pqr}$, rather than the more obvious possibility of eleven dimensional Minkowski space, should be the ground state of this theory.

It will be shown in a separate paper that even when Kaluza–Klein vacuum states are stable classically, they can be destabilized by quantum mechanical tunneling [22]. However, unbroken supersymmetry (plus a technical requirement that the extra dimensions be simply connected; this is not satisfied in the original Kaluza–Klein theory) seems to be a sufficient condition for stability. This is another reason that theories in which $\bar{D}_\mu \varepsilon = 0$ has a solution and there is an unbroken supersymmetry at the energies of compactification would be attractive.

As has been pointed out above, the most serious obstacle to a realistic model of the type considered in this paper is that the fermion quantum numbers do not turn out right. It is conceivable that this problem could be overcome if instead of riemannian geometry one considered geometry with torsion or some generalization of torsion; in such a theory the fermion transformation laws might be different.

How can one obtain torsion in eleven-dimensional supergravity? As has been noted [11], the theory formally contains torsion in the sense that certain fermion bilinears enter, formally, in the way that torsion would appear. Of course, a "torsion" that is bilinear in Fermi fields does not have a classical limit. However, by analogy with QCD, in which $\bar{q}q$ has a vacuum expectation value, one may be willing in supergravity to assume a vacuum expectation value for the "torsion field" $K \sim \bar{\psi}\psi$ (or perhaps for some other bilinears). Perhaps in this way the predictions for fermion quantum numbers can be modified. This possibility is under study.

I wish to acknowledge discussions with V. Bargmann and J. Wolf.

Note added in proof

For a recent discussion of Dirac zero modes in Kaluza–Klein theories, see ref. [23].

References

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys. K1 (1921) 966

[2] O. Klein, Z. Phys. 37 (1926) 895; Arkiv. Mat. Astron. Fys. B 34A (1946); Contribution to 1955 Berne Conf., Helv. Phys. Acta Suppl. IV (1956) 58

[3] A. Einstein and W. Mayer, Preuss. Akad. (1931) p. 541, (1932) p. 130;
A. Einstein and P. Bergmann, Ann. Math. 39 (1938) 683;
A. Einstein, V. Bargmann, and P.G. Bergmann, Theodore von Kármán Anniversary Volume (Pasadena, 1941) p. 212;
O. Veblen, Projektive Relativitats Theorie (Springer, Berlin, 1933);
W. Pauli, Ann. der Phys. 18 (1933) 305; 337;
P. Jordan, Ann. der Physik (1947) 219;
Y. Thirz, C. R. Acad. Sci. 226 (1948) 216

[4] P.G. Bergmann, Introduction to the theory of relativity (Prentice-Hall, New York, 1942);
A. Lichnerowicz, Theories relativistes de la gravitation et de l'electromagnetisme (Masson, Paris, 1955)

[5] J. Rayski, Acta Phys. Pol. 27 (1965) 89;
W. Thirring, *in* 11th Schladming Conf., ed. P. Urban (Springer-Verlag, Wien, New York, 1972)

[6] B. DeWitt, *in* Lectures at 1963 Les Houches School, Relativity, groups, and topology, ed. B. DeWitt and C. DeWitt (New York, Gordon and Breach, 1964), published separately under the title Dynamical theory of groups and fields (New York, Gordon and Breach, 1965);
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Rep. Math. Phys. 1 (1970) 29;
Y.M. Cho, J. Math. Phys. 16 (1975) 2029;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
Y.M. Cho and P.S. Jang, Phys. Rev. D12 (1975) 3789;
L.N. Chang, K.I. Macrae, and F. Mansouri, Phys. Rev. D13 (1976) 235

[7] J. Scherk and J.H. Schwarz, Phys. Lett. 57B (1975) 463;
E. Cremmer and J. Scherk, Nucl. Phys. B103 (1976) 393; B108 (1976) 409

[8] J.-F. Luciani, Nucl. Phys. B135 (1978) 111

[9] L. Palla, Proc. 1978 Tokyo Conf. on High-energy physics, p. 629

[10] N. Manton, Nucl. Phys. B158 (1979) 141

[11] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409;
E. Cremmer and B. Julia, Phys. Lett. 80B (1978) 48; Nucl. Phys. B159 (1979) 141

[12] J. Scherk and J.H. Schwarz, Phys. Lett. 82B (1979) 60; Nucl. Phys. B153 (1979) 61;
E. Cremmer, J. Scherk, and J.H. Schwarz, Phys. Lett. 84B (1979) 83

[13] W. Nahm, Nucl. Phys. B135 (1978) 149

[14] M.T. Grisaru, H.N. Pendleton and P. van Nieuwenhuysen, Phys. Rev. D15 (1977) 996

[15] F.A. Berends, J.W. van Holten, B. deWit and P. van Nieuwenhuizen, Nucl. Phys. B154 (1979) 261; Phys. Lett. 83B (1979) 188; J. Phys. A13 (1980) 1643;
C. Aragone and S. Deser, Phys. Lett. 85B (1979) 161; Phys. Rev. D21 (1980) 352;
K. Johnson and E.C.G. Sudarshan, Ann. of Phys. 13 (1961) 126;
G. Velo and D. Zwanziger, Phys. Rev. 186 (1967) 1337

[16] N. Hitchin, J. Diff. Geom. 9 (1974) 435;
S.T. Yau, Proc. Nat. Acad. Sci. US 74 (1977) 1798;
S. Hawking, Nucl. Phys. B146 (1978) 381

[17] R. Brown and A. Gray, Trans. Amer. Math. Soc. 141 (1969) 465

[18] P. Fayet, Phys. Lett. 69B (1977) 489; 84B (1979) 421;
G.R. Farrar and P. Fayet, Phys. Lett. 76B (1978) 575; 79B (1978) 442; 89B (1980) 191

[19] S. Weinberg, Phys. Lett. 62B (1976) 111

[20] A. Aurilia, H. Nicolai and P.K. Townsend, preprint, CERN Th. 2884 (1980);
P. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233

[21] E. Witten, preprint, Princeton Univ. (April, 1981)

[22] E. Witten, in preparation

[23] W. Mecklenburg, preprint, ICTP (March, 1981)

CHAPTER 6

Cosmology

Where has the fifth dimension gone?

Alan Chodos and Steven Detweiler
J. W. Gibbs Laboratory, Physics Department, Yale University, New Haven, Connecticut 06520

(Received 11 June 1979)

We show that a simple solution to the vacuum field equations of general relativity in $4+1$ space-time dimensions leads to a cosmology which at the present epoch has $3+1$ observable dimensions in which the Einstein-Maxwell equations are obeyed. The large ratio of the electromagnetic to gravitational forces is a consequence of the age of the Universe, in agreement with Dirac's large-number hypothesis.

Over the years there has been a number of attempts[1-7] to construct unified field theories based on five-dimensional space-time, going back to that of Kaluza in 1921. The general scheme is that four dimensions, say $x^0, \ldots, x^3$, are identified with the observed space-time, and the associated ten components of the metric are used to describe gravity. The metric components connecting x^5 to $\{x^0, \ldots, x^3\}$ give four extra degrees of freedom which may be interpreted as the electromagnetic potential. In addition, there is a 15th degree of freedom (essentially g^{55}) which may either be set to one,[1] or else may be allowed to vary,[4,5] thereby introducing a scalar field into the problem.

It is remarkable that, under the assumption that a suitable Killing vector exists, Einstein's field equations in five dimensions look exactly like the Einstein-Maxwell system in four dimensions, with or without an extra scalar field. The great difficulty with this approach is to understand why such a Killing vector should exist, or, more loosely, why the observed Universe is four and not five dimensional. Indeed, the assumption of a Killing vector serves at least partially to dismantle the five-dimensional framework and the hoped-for unification of gravity and electromagnetism is severely compromised, if not entirely lost.

In this paper we discuss a model of a five-dimensional universe which naturally evolves into an effective four-dimensional one, even though all spatial dimensions are treated symmetrically in the field equations and the boundary conditions. As the reader will see, we achieve this at the cost of making some special choices in the solution to the field equations. In the absence of any justification for these choices,[8] we cannot claim to have proven, given our field equations, that the Universe must have evolved along the lines we suggest. Rather, it is our purpose to point out the possibility of such evolution, and to deduce therefrom certain interesting observational consequences.

We let time be a continuous parameter ranging over the real line, and we take the spatial x^i to be periodic; i.e., we assume that in a suitable coordinate system, they can be chosen to have the range

$$0 \leq x^i < L ,$$

where L is some parameter with the dimension of length. In our model, in the absence of matter (which we are temporarily disregarding), the basic equations are then

$$R_{\mu\nu} = 0, \tag{1}$$

where $R_{\mu\nu}$ is the five-dimensional Ricci tensor.

We assume that the Universe has evolved according to the Kasner solution[9] to Eq. (1), which is a particularly simple nontrivial homogeneous vacuum solution. Furthermore, there is reason to believe that many other five-dimensional homogeneous cosmologies will mimic the Kasner solution in much the same way that in four dimensions the mixmaster solutions exhibit Kasner-type behavior.[10] In d spatial dimensions, the Kasner solution has the form

$$ds^2 = -dt^2 + \sum_{i=1}^{d} (t/t_0)^{2p_i} (dx^i)^2, \tag{2}$$

which will solve Eq. (1) provided

$$\sum_{i=1}^{d} p_i = \sum_{i=1}^{d} p_i^2 = 1 . \tag{3}$$

Except in degenerate cases, these relations require at least one of the p_i to be negative. Thus in four dimensions the vacuum Kasner solution is a poor description of the Universe, since it predicts contraction in at least one dimension. Conventionally, this situation is improved[11] by adding matter on the right-hand side of Eq. (1); however, we propose to consider the Kasner solution in five dimensions, in which case the resolution of this difficulty is somewhat different.

In order to guarantee the appearance of isotropy, we take $p_1 = p_2 = p_3 = \frac{1}{2}$, $p_5 = -\frac{1}{2}$. It is possible that these choices may be justified by examining the stability of this solution compared to ones given by other values of the parameters. Then

$$ds^2 = -dt^2 + (t/t_0)[(dx^1)^2 + (dx^2)^2 + (dx^3)^2] + (t_0/t)(dx^5)^2 .$$

At the instant $t = t_0$, the Universe was spatially flat and appeared isotropic, with four space dimensions. The distance around any dimension was L. For $t \ll t_0$, the Universe had essentially only one spatial dimension, whose length approached infinity at the initial singularity $t = 0$. For $t \gg t_0$, the distance around the fifth dimension has shrunk to $(t_0/t)^{1/2}L$, while the other three spatial dimensions have grown to $(t/t_0)^{1/2}L$.

Assuming $(t_0/t)^{1/2}L$ is sufficiently small, i.e., that the Universe is sufficiently old, the fact that the fifth dimension is not observed is a consequence of the dynamics of the evolution of the cosmos, and not of the preordained existence of a Killing vector[1-5] or of spontaneous compactification[12] or some similar mechanism.

We can attempt to estimate the magnitude of $(t_0/t)^{1/2}L$ in a standard way. First we rescale our coordinates so that the Kasner metric takes the form

$$ds^2 = -dt^2 + (t/\tau)(d\vec{x})^2 + (\tau/t)(dx^5)^2 ,$$

where $t/\tau \approx 1$. In so doing, we must remember that our new x^5 satisfies

$$0 \leq x^5 < L(t_0/\tau)^{1/2} .$$

Now we add a small perturbation $h_{\mu 5}$ $(\mu = 0, 1, 2, 3)$ to this metric, where $h_{\mu 5}$ is related to the electromagnetic potential A_μ by

$$\lambda_\mu \equiv h_{\mu 5} = (16\pi G)^{1/2} A_\mu .$$

[This can be derived by demanding that the space-time components of Eq. (1) take the form

$$\tilde{R}_{ab} - \tfrac{1}{2}\tilde{g}_{ab}\tilde{R} = 8\pi G(\tilde{T}^{(\mathrm{em})}_{ab} + T'_{ab}),$$

where $\tilde{R}_{ab}$ is the usual four-dimensional Ricci tensor, $\tilde{T}_{ab}$ is the electromagnetic stress-energy tensor, and T'_{ab} is an extra contribution due to the scalar field.]

Following Souriau,[6] we then consider a quantum field ϕ coupled to this metric via the equation

$$\Box_5 \phi + a\phi = 0 , \tag{4}$$

where $\Box_5$ is the covariant five-dimensional d'Alembert operator. We demand that ϕ be periodic with period $L' = L(t_0/\tau)^{1/2}$ in the x^5 coordinate, and we ignore variations in the quantity t/τ, setting it equal to one. Then, through second order in λ we obtain

$$\Box\phi - (q^2 - a)\phi - 2iq(\lambda\cdot\partial\phi) - iq(\partial\cdot\lambda)\phi - q^2(\lambda\cdot\lambda)\phi - (\lambda_{,5})\cdot\partial\phi + 2iq(\lambda\cdot\lambda_{,5})\phi = 0 , \tag{5}$$

where

$$\phi = \hat{\phi}(x^\mu)e^{iqx^5}, \quad q = 2\pi n/L' .$$

This is the same as the Klein-Gordon equation in the presence of an electromagnetic field (plus some other interactions due to the possible variation of λ in the fifth direction), provided we identify the charge e_n as

$$e_n = \frac{\hbar q_n}{c}(16\pi G)^{1/2} .$$

Thus the basic unit of charge obeys the relationship

$$\frac{e^2}{4\pi\hbar c}\,\frac{c^3}{16\pi^2 G\hbar} = \frac{1}{L'^2} = \frac{\tau}{t_0}\,\frac{1}{L^2} , \tag{6}$$

where τ is a time characteristic of the present age of the Universe and t_0 is the time at which the four dimensions of space were equally large. Putting in the known value of the left-hand side, we conclude that

$$(t_0/\tau)^{1/2}L = 2.38\times 10^{-31}\ \mathrm{cm}. \tag{7}$$

Thus the distance around the fifth dimension is currently very small.

It is interesting that Eq. (6) is in conformity with Dirac's large-number hypothesis[13]: The large value of the ratio of the electromagnetic to gravitational coupling constants is a consequence of the age of the Universe. When t was $\approx t_0$, the two interactions were of approximately the same strength.

It should be noted, however, that Eq. (4) is not satisfactory as it stands as an equation to describe any of the known elementary particles (even in an approximation where spin is neglected). The reason is the $q^2\phi$ term in Eq. (5). With the value of q^2 that follows from Eq. (7), we find that this corresponds to a mass for the particle of 5.22 $\times 10^{20}$ MeV (9.3×10^{-7} g). One way out is to adjust the constant a so that $q^2 - a$ is in the range 0.5 to 5×10^3 MeV. This means fine tuning a to 20 decimal places. It also means that the particle under consideration was a tachyon in the past and will become very massive in the future. Altogether, this seems an unsatisfactory resolution to the problem. Other schemes for coupling matter to our five-dimensional metric are currently under investigation.

While the fifth dimension has been shrinking, the other three spatial dimensions have been expanding, with the radius of the Universe given by

$$R(t)=(t/t_0)^{1/2}L\ .$$

Thus $\dot{R}/R=H=1/2t$.

The quantity H is the observed value of Hubble's constant. The reader may be worried that we have ignored an additional effect, namely, that according to Eq. (6), if we assume

$$\alpha=e^2/4\pi\hbar c$$

is constant over cosmological times (and choosing units so that $c=1$) then the product $G\hbar$ must vary as $1/t$. If this leads to variations in atomic frequencies, then the observed value of H could be different from that predicted above.

Let m be a typical mass in atomic physics (e.g., the mass of electron). Then the ratio of electric to gravitational forces among atomic particles is $\sim e^2/Gm^2$. In atomic units, we have $\hbar=$ const, and therefore $G\propto 1/t$. We also take m to be constant in atomic units. [Strictly speaking, this is an extra assumption that does not follow from Eq. (6), but it is a most natural one.] This gives $e^2/Gm^2 \propto t$, as expected from Dirac's hypothesis. In gravitational units, we have $G=$ const, $\hbar\propto 1/t$, and since $\alpha=$ const we conclude that $e^2\propto 1/t$. Then, in order to have $e^2/Gm^2\propto t$, we conclude that $m\propto 1/t$ in gravitational units.

Now let us examine atomic frequencies. A typical atomic energy level will be given by

$$E=mf(\alpha)$$

and a typical frequency by

$$\nu=(1/h)mf(\alpha)\ .$$

We see that in both atomic and gravitational units the quantity $m/\hbar$ is constant, so that no correction is needed to our value of H, which is determined by measuring the red-shift of atomic spectral lines.[14] However, one must make the added assumption that the usual methods of determining galactic distances continue to hold in our cosmology.

Besides the value of Hubble's constant, we also predict the deceleration parameter:

$$q_0\equiv -R\ddot{R}/\dot{R}^2=1\ .$$

We have chosen to treat the case of five dimensions as the simplest generalization of the usual four. In order to include other interactions besides the gravitational and the electromagnetic in this scheme, it is necessary to generalize our picture to more dimensions. A vacuum Kasner solution of the form Eq. (2) exists in any number of spatial dimensions, provided that Eq. (3) is satisfied. If we choose the three positive powers to have the same value p_1 and let the remaining n powers have the same negative value p_2 we find

$$p_1=\frac{3+(3n^2+6n)^{1/2}}{3(n+3)}\ ,$$

$$p_2=\frac{n-(3n^2+6n)^{1/2}}{n(n+3)}\ ,$$

so that for n large,

$$p_1\approx 1/\sqrt{3}\ ,$$

and therefore

$$H\approx 1/\sqrt{3}\,t$$

and

$$q_0\approx 0.73$$

in that case. Thus, modulo the assumptions mentioned above, we have

$$1/2t\leqslant H\leqslant 1/\sqrt{3}\,t$$

and

$$0.73\leqslant q_0\leqslant 1.$$

In this paper we have attempted to convince the reader of the possibility that extra dimensions of space, which have appeared for technical reasons in the literature from time to time, may possess a hitherto unsuspected historical reality.

One of us (A.C.) was supported in part by the U.S. Department of Energy under Contract No. EY-76-C-02-3075. The other (S.D.) was supported in part by NSF Grant No. PHY76-82353.

---

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. LIV, 966 (1921).

[2] O. Klein, Z. Phys. 37, 875 (1926); Nature (London) 118, 516 (1926).

[3] A. Einstein and P. Bergmann, Ann. Math. 39, 683 (1938).

[4] P. Bergmann, Ann. Math. 49, 255 (1948).

[5] Y. Thiry, J. Math. Pure Appl. 30, 275 (1951).

[6] J.-M. Souriau, Nuovo Cimento 30, 565 (1963).

[7] W. Thirring, Acta Phys. Austriaca Suppl. IX, 256 (1972).

[8] In a larger sense, one might raise the question of whether it is ever possible to justify one's choice of solution when one is talking about the Universe as a whole; however, a discussion of this point is beyond the scope of this work.

[9] E. Kasner, Am. J. Math. 48, 217 (1921).

[10] C. W. Misner, Phys. Rev. Lett. 22, 1071 (1969).

[11] E. Schücking and O. Heckmann, in *Onzième Conseil de Physique Solvay* (Éditions Stoops, Brussels, 1958), p. 149.

[12] E. Cremmer and J. Scherk, Nucl. Phys. B103, 399 (1976); J. Scherk and J. H. Schwarz, Phys. Lett. 82B, 60 (1979).

[13] P. A. M. Dirac, Proc. R. Soc. London A165, 199 (1938); A365, 19 (1979).

[14] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 441.

KALUZA-KLEIN COSMOLOGIES*

Peter G.O. FREUND

The Enrico Fermi Institute, and the Department of Physics, University of Chicago, Chicago, Illinois 60637

Received 22 June 1982

In generalized Kaluza-Klein theories the scale set by the size of the extra space-dimensions is close to the grand unification scale of supersymmetric GUT's with minimal number of Higgs supermultiplets. In view of this observation, we explore cosmologies in which the "effective" dimensionality of space depends on time. Such cosmologies are studied in higher-dimensional Jordan-Brans-Dicke theories, and in 10- and 11-dimensional supergravity. The preferential expansion of three space-like dimensions is noted in the latter theory. Cosmology in pure higher-dimensional Einstein theory, where there is no preferential expansion of three space-like dimensions, has been discussed by Chodos and Detweiler.

1. Kaluza-Klein scale and cosmology

In unified theories obtained by dimensional reduction [1] from higher-dimensional gravity or supergravity theories, the size of the extra (compactified) dimensions is determined in terms of Newton's constant G and of the grand unified fine structure constant α

$$l = 4\pi(G/\alpha)^{1/2}. \tag{1.1}$$

Typically $\alpha \approx 0.01$, so that l exceeds the Planck length $(\sqrt{G})$ by two orders of magnitude and is comparable [2] to the grand unification scale of supersymmetric theories with the minimal number of Higgs supermultiplets. Thus, in such generalized Kaluza-Klein theories, at the grand unification scale the world manifold has more than four dimensions and an "effective" four-dimensional description becomes inapplicable [2]. This increase in dimensionality sets in *before* quantum gravity effects become relevant.

In view of all this, it becomes interesting to study cosmologies based on a higher-dimensional theory, which at "later" times become effectively four-dimensional, in that the extra dimensions become inobservably small by either contracting or expanding much less than the physically observable dimensions. A first step in

---

* Work supported in part by the NSF: PHY-79-23669.

this direction was taken in an interesting paper by Chodos and Detweiler [3]. They studied Kasner type solutions [4] of a pure gravitational theory in $4 + N$ dimensions. Aside from stability problems, there is no way in such an approach to account for a preferential compactification of a certain number of dimensions. In higher-dimensional supergravity theories, supersymmetry *requires* the presence of specified matter fields, and it was remarked that these fields may hold the clue to why precisely three space dimensions stay large [5]. Here we shall present cosmological solutions for $N = 1$ supergravity in 10 and 11 dimensions [6, 7]. Without exciting the Bose matter fields the eventual number of large dimensions would remain arbitrary as in ref. [3]. Once one excites these fields there are cosmological solutions that preferentially expand three of the space dimensions. Given that we do not, in these solutions include excitations of the Fermi fields the Fermi matter-dominated era is not covered, and our discussion is presumably restricted to an era close to the "dimensional transition", when three space-like dimensions become differentiated from the others.

As a rule, our solutions contain new parameters, the time at which all space dimensions have comparable size, and the actual common size of these dimensions. If the extra dimensions contract, then a scale considerably larger than the present-day Kaluza-Klein scale of eq. (1) can appear in the theory. It is amusing to note that such scales have been recently considered [8] for apparently very different reasons.

2. Description of geometries

We start with a space-time manifold with one time-like and $d + D$ spacelike dimensions. For cosmological purposes we restrict our attention to metrics of the form

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & R_d^2(t)\tilde{g}_{mn}(x^P) & 0 \\ 0 & 0 & R_D^2(t)\tilde{g}_{MN}(x^P) \end{pmatrix} \tag{2.1}$$

Here $\mu, \nu, \ldots = 0, 1, \ldots, d + D$; $m, n, \ldots = 1, \ldots, d$; $M, N, \ldots = d + 1, d + 2, \ldots d + D$; $\tilde{g}_{mn}(x^P)$ and $\tilde{g}_{MN}(x^P)$ are the metrics of maximally symmetric d- and D-dimensional spaces respectively and R_d, R_D the corresponding time-dependent cosmological scale factors. It should be obvious that this ansatz is restrictive. There is no a priori reason for maximal symmetry in the extra D dimensions. The ansatz (2.1) rules out in particular manifolds such as $\mathrm{M}_D = CP_2 \times S^2 \times S^1$ considered by Witten [9]. Ultimately this and other not maximally symmetric cases will have to be studied, but here we limit ourselves to the maximally symmetric case for reasons of simplicity.

Throughout this paper we use the notations and definitions from Weinberg's book [10]. The nonvanishing components of the Christoffel connection for the metric (2.1) are then

$$\Gamma^0_{mn} = \frac{\dot R_d}{R_d} g_{mn}, \qquad \Gamma^0_{MN} = \frac{\dot R_D}{R_D} g_{MN}, \qquad \Gamma^l_{m0} = \frac{\dot R_d}{R_d} \delta^l_m, \qquad \Gamma^L_{M0} = \frac{\dot R_D}{R_D} \delta^L_M,$$

$$\Gamma^l_{mn} = \tilde\Gamma^l_{mn}, \qquad \Gamma^L_{MN} = \tilde\Gamma^L_{MN}, \tag{2.2}$$

where $\tilde\Gamma^l_{mn}$ and $\tilde\Gamma^L_{MN}$ are the Christoffel connections for the maximally symmetric metrics $\tilde g_{mn}$ and $\tilde g_{MN}$. The nonvanishing components of the Ricci tensor $R_{\mu\nu}$ are

$$R_{00} = d\frac{\dot R_d}{R_d} + D\frac{\dot R_D}{R_D},$$

$$R_{mn} = -\left(\frac{2k_d}{R_d^2} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot R_d}{R_d}\right) + \left(d\frac{\dot R_d}{R} + D\frac{\dot R_D}{R_D}\right)\frac{\dot R_d}{R_d}\right) g_{mn},$$

$$R_{MN} = -\left(\frac{2k_D}{R_D^2} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot R_D}{R_D}\right) + \left(d\frac{\dot R_d}{R_d} + D\frac{\dot R_D}{R_D}\right)\frac{\dot R_D}{R_D}\right) g_{MN}. \tag{2.3}$$

Here k_d and k_D are the curvature constants of M_d and M_D. Positive, negative values of k_d (or k_D) correspond to the sphere and pseudosphere respectively. Vanishing k_d (k_D) corresponds to a flat M_d (M_D).

3. Jordan-Brans-Dicke theory in *N* Dimensions

We shall encounter below (in 10-dimensional supergravity) higher-dimensional variants of the Jordan-Brans-Dicke (JBD) theory [11]. Therefore we present here some useful results about such theories. In the absence of matter fields JBD theory in N dimensions starts from the action

$$S = -\frac{1}{16\pi}\int \mathrm{d}^N x\sqrt{-g}\left(\phi R + \omega\frac{\nabla_\mu\phi\,\nabla^\mu\phi}{\phi}\right). \tag{3.1}$$

The variational principle

$$\delta S = 0 \tag{3.2}$$

yields the field equations

$$\Box\phi = 0 \tag{3.3a}$$

and

$$G_{\mu\nu} = T_{\mu\nu}(\omega, \phi), \tag{3.3b}$$

where

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R, \tag{3.3c}$$

$$T_{\mu\nu}(\omega, \phi) = -\left[\frac{\omega}{\phi^2}\left(\nabla_\mu \phi\, \nabla_\nu \phi - \tfrac{1}{2} g_{\mu\nu} g^{\alpha\beta} \nabla_\alpha \phi\, \nabla_\beta \phi\right) + \frac{1}{\phi}\left(\nabla_\nu \nabla_\mu \phi - g_{\mu\nu} \Box \phi\right)\right]. \tag{3.3d}$$

A Weyl transformation

$$g_{\mu\nu} \to \hat{g}_{\mu\nu} = \Omega^2(x) g_{\mu\nu}, \qquad \phi \to \hat{\phi} = \phi \tag{3.4}$$

with scale function

$$\Omega(x) = \phi(x)^{-1/(N-2)} \tag{3.5}$$

transforms the action (3.1) into

$$\hat{S} = -\frac{1}{16\pi} \int \mathrm{d}^N x \sqrt{-g} \left(R + \left(\omega + \frac{N-1}{N-2}\right) \nabla_\mu \chi\, \nabla^\mu \chi\right) \tag{3.6}$$

with $\chi = \ln \phi$. The weak energy condition [12] requires the constant multiplying $\nabla_\mu \chi\, \nabla^\mu \chi$ to be non-negative, whence the condition

$$\omega \geqslant \omega_{\mathrm{w}} \equiv -\frac{N-1}{N-2}. \tag{3.7}$$

The limiting case

$$\omega = \omega_{\mathrm{w}} \tag{3.8}$$

corresponds to invariance under the generalized Weyl transformations

$$g_{\mu\nu} \to \Omega^2(x) g_{\mu\nu}, \qquad \phi \to (\Omega(x))^{2-N} \phi. \tag{3.9}$$

Had we also included matter fields with appropriate scaling properties, the field equations (3.3) would be replaced by N-dimensional versions of the Brans-Dicke equations

$$\phi = 4\pi\lambda T_{\mathrm{M}}$$

$$G_{\mu\nu} = T_{\mu\nu}(\omega, \phi) = \frac{8\pi}{\phi} T_{\mathrm{M}_{\mu\nu}} \tag{3.10a}$$

with

$$\lambda = \frac{2}{N-2} \frac{1}{\omega - \omega_{\mathrm{w}}}, \tag{3.10b}$$

$T_{\mathrm{M}_{\mu\nu}}$, T_{M} the matter energy-momentum tensor and its trace, and ω_{w} given by eq. (3.7). The so-called Mach principle constraint [11]

$$\lambda > 0 \tag{3.11}$$

is thus seen to yield nothing beyond the weak energy condition (3.7). In their four-dimensional work [11] Brans and Dicke require a positive ω. Beyond the weak energy condition and/or Mach's principle, this also assumes that $|\omega|$ is large enough for the ratio $(2\omega + 3)/(2\omega + 4)$, which measures the departure from Einstein theory, to be sufficiently close to 1 so as not to contradict experimental evidence. In N dimensions where no experimental evidence is available, all we can reasonably require is the energy condition (3.7).

JBD fields are known to appear in dimensional reduction [1, 11]. Suppose D of the original $d + D$ space-like dimensions are compactified into a "small" D-manifold M_D. The square root of the determinant of the M_D metric then functions as a JBD field in the reduced $d + 1$ dimensional theory and the corresponding value of ω is

$$\omega = -1 + \frac{1}{D}, \tag{3.12}$$

which indeed obeys the energy condition (3.7) no matter what values d and D take. In particular for $D = 1$ when only one dimension is compactified, as in Jordan's original work, we have $\omega = 0$.

4. Cosmological solutions in higher-dimensional Jordan-Brans-Dicke theory

As a last preliminary to cosmological solutions in higher-dimensional supergravity theories, we now explore cosmological solutions for higher-dimensional JBD theory.

We assume a metric of the form (2.1) and a JBD field $\phi(t)$ depending only on the time variable. For

$$k_d = k_D = 0 \tag{4.1}$$

in eqs. (2.3), the general solution of the field equations (3.3) is

$$R_d(t)=r_d(t+t_0)^{\alpha_d}, \qquad R_D(t)=r_D(t+t_0)^{\alpha_D}, \qquad \phi(t)=\varphi(t+t_0)^{\gamma}, \tag{4.2a}$$

where $r_d, r_D, \varphi, t_0, \alpha_d, \alpha_D, \gamma$ are constants obeying the constraints

$$\begin{gathered}\gamma+d\alpha_d+D\alpha_D=1, \\ (\gamma-1)(\gamma+1-\alpha_d-\alpha_D)-(d+D)\alpha_d\alpha_D+\omega\gamma^2=0.\end{gathered} \tag{4.2b}$$

The d- and D-dimensional space-like manifolds M_d and M_D are both flat [eq. (4.1)]. How they expand or contract is governed by eqs. (4.2) for given ω. As a simple example consider the case of a "steady-state" cosmology defined by the condition that, say, d dimensions expand and D contract, in such a way that the volume

$$(R_d)^d(R_D)^D=\text{constant} \tag{4.3a}$$

or

$$d\alpha_d+D\alpha_D=0. \tag{4.3b}$$

In this case the first constraint (4.2b) fixes

$$\gamma=1 \tag{4.3c}$$

and the second constraint (4.2b) along with eqs. (4.3b) and (4.3c) then give

$$\alpha_d=\left(-\frac{D}{d}\frac{\omega}{D+d}\right)^{1/2}, \qquad \alpha_D=-\left(-\frac{d}{D}\frac{\omega}{D+d}\right)^{1/2}. \tag{4.4}$$

This requires $\omega<0$ which, as we saw in sect. 3, is allowed. We note that for $\gamma=0$ eqs. (4.2) reduce to

$$\begin{gathered}d\alpha+D\beta=1, \\ d\alpha^2+D\beta^2=1,\end{gathered}$$

and the solution is of the Kasner type [3, 4]. From these examples we see that there exist solutions with d space-like dimensions expanding and D contracting.

In the absence of extra constraints such as the "steady state" condition (4.3a), there are whole families of solutions for a given ω. This is not the only ambiguity. At this level there is nothing to fix the JBD parameter ω or the dimensionalities d and D

of the space-like manifolds M_d and M_D. Therefore we consider higher-dimensional supergravities. Of these, there is only a finite sample and even then only two cases are relevant: $1 + d + D = 10$, and $1 + d + D = 11$. In the 11-dimensional case there is no JBD field at all, whereas in the ten dimensional case there is such a field. Since the 10-dimensional theory can be obtained from the eleven-dimensional one by dimensional reduction one has $\omega = 0$ as dictated by eq. (3.12). Moreover, in either case one has antisymmetric tensor matter fields required by supersymmetry. These fields can then be used to fix d and D individually as in ref. [5].

5. Cosmological solutions in eleven-dimensional supergravity

We first treat maximal eleven-dimensional supergravity [6]. In its Bose sector this theory involves gravity and an antisymmetric tensor matter field $A_{\mu\nu\rho}$ ($\mu, \nu, \rho = 0, 1, 2, \ldots, 10$). Following ref. [5] we use this field $A_{\mu\nu\rho}$ to preferentially split a space-like three-dimensional manifold M_3 from the remaining seven-dimensional space M_7. In other words, for the geometry described in sect. 2 we get $d = 3$, $D = 7$ (and of course one time dimension). This is done by requiring $A_{\mu\nu\rho}$ to live (i.e., to be a maximally form-invariant tensor) on M_3

$$A_{\mu\nu\rho} = \sqrt{g_3}\, \varepsilon_{\mu\nu\rho} A(t). \tag{5.1a}$$

Here $A(t)$ is a function of time, g_3 is the metric determinant of M_3, and $\varepsilon_{\mu\nu\rho}$ the Levi-Cività symbol when all its three indices correspond to M_3, zero otherwise [5]. Define the field strengths

$$F_{\mu\nu\rho\sigma} = \partial_\mu A_{\nu\rho\sigma} + \partial_\nu A_{\rho\sigma\mu} + \partial_\rho A_{\sigma\mu\nu} + \partial_\sigma A_{\mu\nu\rho}. \tag{5.2}$$

For the corresponding contravariant components $F^{\mu\nu\rho\sigma}$ the ansatz (5.1a) yields the expression

$$F^{\mu\nu\rho\sigma} = \frac{\varepsilon^{\mu\nu\rho\sigma}}{\sqrt{-g_4}} F(t), \qquad F(t) = -\dot{A}(t), \tag{5.1b}$$

where g_4 is the metric determinant of the four-dimensional space-time M_4 resulting from the addition of time to M_3 and $\varepsilon^{\mu\nu\rho\sigma}$ the corresponding Levi-Cività symbol [5]. With the ansatz (5.1) the term trilinear in $A_{\mu\nu\rho}$ and its derivatives in the eleven dimensional supergravity action S_{11} vanishes. The Bose part of S_{11} becomes

$$S_{11} = -\frac{1}{16\pi G_{11}} \int \mathrm{d}^{11}x \sqrt{-g_{11}} \left(R + \tfrac{1}{3}\pi G_{11} F_{\mu\nu\rho\sigma} F^{\mu\nu\rho\sigma} \right) \tag{5.2}$$

with G_{11} an 11-dimensional gravitational constant. The $A_{\mu\nu\rho}$ field equations (still with the ansatz (5.1)) are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left((R_7)^7 F\right) = 0, \tag{5.3}$$

where R_7 and (R_3) are defined as in eq. (2.1) with $d = 3$, $D = 7$. The Bianchi identities for $F_{\mu\nu\rho\sigma}$ yield nothing new. The Einstein equations are

$$G_{\mu\nu} = -8\pi G_{11}\theta_{\mu\nu}, \tag{5.4}$$

where

$$\theta_{\mu\nu} = -\tfrac{1}{2}\varepsilon F^2 g_{\mu\nu},$$

$$\varepsilon = \begin{cases} +1 \text{ on } \mathrm{M}_4 \\ -1 \text{ on } \mathrm{M}_7 \end{cases} \tag{5.5}$$

and $G_{\mu\nu}$ is given by eq. (3.3c). Using equations (2.3) with $d = 3$ and $D = 7$ and eqs. (5.3) we find the following solutions.

First, there is a solution with both M_3 and M_7 in cosmological expansion, M_3 expanding much faster than M_7. This solution is

$$R_3(t) = r_3(t + t_0), \qquad R_7(t) = r_7(t + t_0)^{1/7},$$
$$F(t) = \tfrac{3}{2}(14\pi G_{11})^{-1/2}(t + t_0)^{-1}, \tag{5.6}$$
$$k_7 = 0, \qquad |k_3| = \tfrac{27}{14}r_3, \qquad \frac{k_3}{|k_3|} = -1$$

with r_3, r_7 positive integration constants. M_7 is thus flat (i.e., a locally, though not globally, maximally symmetric torus) and M_3 is a pseudosphere. After three spatial dimensions have grown much larger than the remaining seven, G_{11} divided by the volume of M_7 functions as an effective four-dimensional gravitational constant G_4. Since from equation (5.6) $R_7 \sim t^{1/7}$, we have $G_4 \sim t^{-1}$ as proposed by Dirac [13].

For $k_7 > 0$ we have a solution of the type considered in reference 5, with $R_7(t)$ and $F(t)$ independent of time and R_3 evolving as an anti-de-Sitter universe ($k_3 < 0$)

$$R_3(t) = R_3(0)\cos \alpha t \tag{5.7}$$

with α determined in terms of G_{11} and the constant value of F.

We recall that our solutions are valid only in the very early universe, as explained in sect. 1.

6. Cosmological solutions of ten-dimensional $N = 1$ supergravity

Ten-dimensional supergravity has been advocated as a fundamental theory by Scherk and Schwarz [14] on the basis of its connection with fermionic string theory. In this case along with gravity the Bose sector contains a JBD field and a Kalb-Ramond field $A_{\mu\nu}$ [15, 7]. The ω parameter of the JBD field vanishes since the theory can be obtained by dimensional reduction from eleven dimensions so that formula (3.12) with $D = 1$ applies. Call $F_{\mu\nu\rho}$ the Kalb-Ramond field strengths. The Bose part of the action is [7]

$$S = -\frac{1}{16\pi}\int \mathrm{d}^{10}x\left(\phi R + \frac{1}{24\phi}F_{\mu\nu\rho}F^{\mu\nu\rho}\right). \tag{6.1}$$

As before we assume a metric of form (2.1) with $d = 3$, $D = 6$ corresponding to the space-like part of the world manifold being the product of a 3-dimensional maximally symmetric space M_3 and of a 6-dimensional maximally symmetric space M_6. Again we require $F^{\mu\nu\rho}$ to live on M_3

$$F^{\mu\nu\rho} = \frac{\varepsilon^{\mu\nu\rho}}{\sqrt{g_3}}F, \tag{6.2}$$

where g_3 is the metric determinant of M_3, $\varepsilon^{\mu\nu\rho}$ is the Levi-Cività tensor when all its indices correspond to M_3 and zero otherwise and F a function of time.

The ϕ-field equation is

$$R - \frac{1}{24\phi^2}F_{\mu\nu\rho}F^{\mu\nu\rho} = 0. \tag{6.3}$$

The Kalb-Ramond field equation becomes an identity in view of the ansatz (6.2). Yet the Bianchi identities for $F_{\mu\nu\rho}$ require

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[(R_3)^3 F\right] = 0. \tag{6.4}$$

Finally, the Einstein equations are

$$G_{\mu\nu} = \frac{1}{\phi}\theta_{\mu\nu} + T_{\mu\nu}(0, \phi) \tag{6.5a}$$

with $T_{\mu\nu}(0, \phi)$ given by eq. (3.3d), and

$$\theta_{\mu\nu} = -\frac{1}{8\phi}\left(F_{\mu\alpha\beta}F_{\nu}{}^{\alpha\beta} - \tfrac{1}{6}g_{\mu\nu}F_{\alpha\beta\gamma}F^{\alpha\beta\gamma}\right). \tag{6.5b}$$

In view of the ansatz (6.2)

$$\theta_{\mu\nu} = -\varepsilon \frac{F^2}{8\phi} g_{\mu\nu},$$

$$\varepsilon = \begin{cases} +1 & \mu, \nu \text{ on } \mathrm{M}_3 \\ -1 & \text{otherwise.} \end{cases} \tag{6.5c}$$

The Einstein equations can then be cast in the simplified form

$$R_{\mu\nu} = -\frac{F^2}{24\phi^2}(1+3\varepsilon) g_{\mu\nu} - \frac{\nabla_\nu \nabla_\mu \phi}{\phi} . \tag{6.6}$$

If F/ϕ is time independent then the first term on the right-hand side reduces to a cosmological term, albeit with different values of the cosmological constant on and off M_3 on account of the appearance of ε (this is similar to the case considered in ref. [5]). A solution of this type exists with

$$R_3(t) = R_3(0), \qquad \phi(t) = \phi(0), \qquad F(t) = F(0) \tag{6.7a}$$

all time independent and

$$R_6(t) = R_6(0) \cos \alpha t \tag{6.7b}$$

with

$$\alpha^2 = \left(\frac{F(0)}{6\phi(0)} \right)^2 = -\frac{2k_6}{5(R_6(0))^2} = \frac{k_3}{6(R_3(0))^2} . \tag{6.7c}$$

This requires $k_6 < 0$ so that M_6 is not a compact sphere. With $k_3 = 0$ and F/ϕ time-dependent we can find another solution

$$R_3(t) = a_3 t^{1/4}, \qquad R_6(t) = a_6 t, \qquad \phi(t) = a_\phi t^{1/4},$$

$$F(t) = a_F t^{-3/4}, \qquad k_6/a_6^2 = -23/8,$$

and again $k_6 < 0$. To boot, R_6 increases faster than R_3.

Unlike the 11-dimensional case, the 10-dimensional does not seem to preferentially expand three space dimensions. It should be noted that all JBD cosmological solutions (4.2b) with $\omega = 0$ and $D + d = 9$ are also solutions of $N = 1$ supergravity in 10 dimensions corresponding to $F_{\mu\nu\rho} = 0$. Among these there are also solutions expanding 3 space dimensions more than the remaining 6 but there is no reason to prefer these over solutions where some other number of space dimensions expands.

It remains to be seen whether cosmological solutions with preferential expansion of three dimensions exist in 10-dimensional $N = 1$ supergravity.

7. Discussion of results

We have found a variety of ways in which a universe with $3 + N$ comparably sized space dimensions at one era can evolve to a universe in which 3 space dimensions are much "larger" than the remaining N. To account for the number of large dimensions we invoked the special matter fields dictated by supersymmetry in a manner similar to ref. [5]. These solutions did not involve the Fermi fields, and as such should be of an applicability limited to some time-interval around the era where the universe changes over from $3 + N$ to 3 space dimensions. Yet even at this level new scales appear: the time t_s at which the small and large dimensions have comparable size, and the actual sizes $l(t_s)$ of the dimensions at time t_s. Depending on the details of the evolution these new scales can considerably exceed both the Planck scale and the present day Kaluza-Klein scale (1.1). Moreover, even the strength of gravity at time t_s can be appreciably larger than that corresponding to the Planck scale. It thus appears that cosmology of the very early universe may differ considerably from the usually contemplated "scenarios".

References

[1] Th. Kaluza, Sitzungsber, Preuss. Akad. Wiss. Phys. Math. Kl. 966 (1921);
O. Klein, Z. Phys. 37 (1926) 895;
B. de Witt, Dynamical theories of groups and fields (Gordon and Breach, New York, 1965) p. 139;
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Rep. Math. Phys. 1 (1970) 29;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711
[2] P.G.O. Freund, Enrico Fermi Institute preprint EFI 82/23 (1982)
[3] A. Chodos and S. Detweiler, Phys. Rev. D21 (1980) 2167
[4] E. Kasner, Am. J. Math. 48 (1921) 217
[5] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233
[6] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409
[7] A.H. Chamseddine, Nucl. Phys. B185 (1981) 403
[8] E. Witten, ICTP Trieste preprint IC/81/106;
S. Dimopoulos, private communication (1982)
[9] E. Witten, Nucl. Phys. B186 (1981) 412
[10] S. Weinberg, Gravitation and cosmology (Wiley, N.Y., 1972)
[11] P. Jordan, Schwerkraft und Weltall (Vieweg, Braunschweig, 1955);
Z. Physik 157 (1959) 112;
C. Brans and R.H. Dicke, Phys. Rev. 124 (1961) 925
[12] S. Hawking and G.F.R. Ellis, The large scale structure of spacetime (Cambridge Univ. Press, Cambridge, 1973) p. 40.
[13] P.A.M. Dirac, Proc. Roy. Soc. (London) A165 (1938) 199
[14] J. Scherk and J.H. Schwarz, Nucl. Phys. B81 (1974) 118;
M.B. Green, J.H. Schwarz and L. Brink, Nucl. Phys. B198 (1982) 474
[15] F. Gliozzi, J. Scherk and D. Olive, Nucl. Phys. B122 (1977) 253

Entropy from Extra Dimensions

Enrique Alvarez[a]
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

and

M. Belén Gavela[b]
Department of Physics, Brandeis University, Waltham, Massachusetts 02254

(Received 25 May 1983)

The phenomenon of "dynamical compactification" in a universe with $4+n_c$ dimensions is studied, and it is found that if the whole process is adiabatic, entropy is pumped into the effective four-dimensional universe. Some cosmological consequences of this fact are discussed.

PACS numbers: 98.80.Bp, 04.50.+h, 11.30.Pb

It is possible that space-time has more than four dimensions and that the extra ones have so far escaped detection.[1] The earliest suggestion in this respect is Kaluza and Klein's idea that the topology of an extra dimension is a circle S^1 of a very small radius, obtaining in this way a unified theory of gravitation and electromagnetism. The modern version[2] of this idea, called spontaneous compactification, proves that a manifold of the type $V_4 \times C_{n_c}$, where V_4 is the ordinary four-dimensional space-time and C_{n_c} is a compact space, is a static solution of Einstein's equations in $D \equiv 4+n_c$ dimensions and, as such, is as good a candidate for the vacuum of the quantum theory as D-dimensional Minkowski space.

In the event that the space-time has indeed extra dimensions, it is not sufficient to have proved that spontaneous compactification occurs; cosmological dynamics of the extra dimensions must also be studied,[3] explaining why the universe appears as four dimensional now, and what are the consequences of the disappearance, for all practical purposes, of the extra dimensions, while the known ones are well described by a Friedman-Robertson-Walker type of universe.

The main point to be made in this Letter is that it is possible to imagine in the framework of classical general relativity, a scenario in which the extra dimensions keep shrinking while the normal ones expand, and that if the whole process is adiabatic in the D-dimensional sense, a large amount of entropy can be produced. The production of entropy by this process can perhaps be viewed as somewhat similar to the sudden smashing of a D-dimensional lattice into a $D-1$ dimensional one, with the consequent increase in disorder. This process cannot go on forever, though. After the size of the extra dimensions becomes smaller than a critical size to be specified later, entropy production virtually stops and the expansion proceeds in the standard adiabatic way.

The entropy thus produced could eventually, depending upon the details of the solution, be as big as necessary to solve the horizon and flatness problems, providing hypothetical alternatives to the usual inflationary scenario.[4] Another possibility is that this entropy spreads out as uniform background radiation.

We will assume, for simplicity, that the metric has the Bianchi type-I form

$$ds^2 = dt^2 - R^2\delta_{ij}dx^i dx^j - A^2\delta_{ab}dy^a dy^b, \qquad (1)$$

where $i,j=1,2,3$ (the ordinary space dimensions) and $a,b=4,\ldots,D$ (the compact dimensions).[5] The compact dimensions are "periodic" in the sense of being restricted to $-L \leq y_i \leq L$, where L is the constant fixing the scale of length in the extra dimensions. Here, $R(t)$ and $A(t)$ are functions of time only, to be determined from Einstein's equations ($m,n=0,\ldots,D$)

$$R_{mn} - \tfrac{1}{2}(R-2\lambda)g_{mn} = -8\pi G T_{mn}, \qquad (2)$$

where the energy-momentum tensor is assumed to be of the perfect-fluid form

$$T_{mn} = (\rho+p)u_m u_n - p g_{mn}. \qquad (3)$$

The integrability condition $\nabla_m T^{mn}=0$ implies a constraint among the functions $\rho(t)$, $p(t)$, $R(t)$, and $A(t)$:

$$\dot{\rho} + (\rho+p)(3\dot{R}/R + n_c\dot{A}/A) = 0 \qquad (4)$$

where $\dot{\rho} \equiv d\rho/dt$, etc., and the pressure p is given in terms of ρ by an equation of state [$p=0$ for "dust," $p=(D-1)^{-1}\rho$ for "radiation," etc.].

Let us study now what are the effects of having n_c dimensions contracting. We shall assume that this contraction happens in an adiabatic way, i.e., that

$$\nabla_m S_D{}^m = 0, \tag{5}$$

where $S_D{}^m$, the entropy D-vector, is given in terms of the mean four-velocity by $S_D{}^m = S_D \cdot u^m$. The entropy density will then obey the relation

$$S_D R^3 A^{n_c} \equiv X = \text{const}, \tag{6}$$

which implies that the total entropy in a unit of comoving volume is constant.

Let us first assume that the D-dimensional universe is uniformly filled with radiation (a similar analysis can be done for dust or other matter). Since the momenta associated with the extra dimensions will take on only discrete values, $q_i = 2\pi m_i / AL$, then the entropy density for this D-dimensional photon gas will be (in the local inertial frame)

$$S_D(t) = -\frac{n_c + 2}{(2\pi)^3 (LA)^{n_c}}\, k \sum_{\vec{m}} \int d^3p \{f_D \ln f_D - (1+f_D)\ln(1+f_D)\}, \tag{7}$$

where f_D is Planck's distribution function, which in this frame reads

$$f_D = \left(\exp\{\beta[p^2 + (2\pi/LA)^2 \vec{m}^2]^{1/2}\} - 1\right)^{-1}. \tag{8}$$

In order to study the thermal evolution of the radiation gas, it is useful to consider two extreme periods: epoch I, when the extra dimensions are of an enormous size, so that $AL \gg \beta$, and epoch II, in which $AL \ll \beta$.

At epoch I, the discrete sum can be replaced by an integral, so that

$$S_D{}^{(\mathrm{I})} = c_D \beta^{-n_c - 3}, \tag{9}$$

where

$$c_D = \frac{k(n_c+2)!\,\varphi(n_c+4)}{\pi^{n_c/2} 2^{n_c+2} \Gamma((n_c + 3/2))}.$$

For a radiation-dominated universe ($\rho \sim R^{-3(4+n_c)/(3+n_c)} A^{-n_c(4+n_c)/(3+n_c)}$) the D-dimensional adiabaticity condition implies

$$T^{(\mathrm{I})} \sim R^{-3/(3+n_c)} A^{-n_c(3+n_c)}. \tag{10}$$

All the intrinsic D-dimensional quantities must now be related to some four-dimensional ones. It should be stressed that this is an artificial procedure during epoch I, and it is only justified by the fact that we know that after some time (namely, during epoch II) the four-dimensional universe will be singled out.

The D-dimensional distribution function is normalized via the particle density; i.e., the number of particles in a $D-1$ dimensional hypersurface V_{3+n_c} is given by

$$N(V_{3+n_c}) = \int f_D(p, q, t) R^3 d^3x\, A^{n_c} d^{n_c}y\, R^3 d^3p\, A^{n_c} d^{n_c}q. \tag{11}$$

This number can also be obtained in a formal way using $V_{3+n_c} = V_x \times V_y$, where V_x is a three-dimensional hypersurface and V_y an n_c-dimensional one. For an observer not sensitive to the extra dimensions, the probability of finding a particle in V_x with momentum between p and $p + dp$ at time t is intuitively given by integrating out f_D over the extra degrees of freedom, namely y and q. In fact, the equivalent four-dimensional distribution function $f_4(p, t)$ can be defined in a natural way by

$$f_4(p, t) \equiv \frac{\int f_D A^{n_c} d^{n_c}q}{\int A^{n_c} d^{n_c}q} \equiv \frac{\int f_D\, d^{n_c}q}{V(q)}, \tag{12}$$

so that

$$\int f_4 R^3 d^3x\, R^3 d^3p = \frac{N(V_{3+n_c})}{A^{n_c} V(q) \int A^{n_c} d^{n_c}y} \equiv \frac{N(V_{3+n_c})}{A^{n_c} V(q) A^{n_c} V(y)}.$$

Both $V(q)$ and $V(y)$ are formally infinite, reflecting the fact that we have normalized in D dimensions to a finite quantity. They will eventually disappear though in all ratios of physical quantities computed with f_4. In the second epoch (when $A \to 0$), we recover the natural result

$$f_4(p, t) = f_D(p, A = 0, t).$$

Explicit computation of f_4 in epoch I, using (12) and (8) gives

$$f_4 = V(q)^{-1}\pi^{(n_c-1)/2}2^{(n_c+1)/2}A^{-n_c}\beta^{(1-n_c)/2}(Rp)^{(n_c+1)/2}\sum_{m=1}^{\infty} m^{(1-n_c)/2}K_{(1+n_c)/2}(\beta m R p), \tag{13}$$

whose infrared limit is

$$f_4 \sim V(q)^{-1}2^{n_c}\pi^{(n_c-1)/2}[(n_c-1)/2]!(\beta A)^{-n_c}\varphi(n_c), \tag{14}$$

and which in the ultraviolet behaves as

$$f_4 \sim V(q)^{-1}2^{n_c/2}\pi^{n_c/2}A^{-n_c}\beta^{-n_c/2}(Rp)^{n_c/2}e^{-\beta R p}. \tag{15}$$

These limit behaviors are similar to the ones for the four-dimensional Planck function; the main differences are that the zero in the ultraviolet gets some power corrections, and that the infrared behavior is regulated by an effective cutoff.

In view of the complicated exact expression for f_4, it seems only natural to represent its physical effects by a Planck function with an effective temperature determined in such a way that the mean energy of this fictitious Planck gas is the same as the one stemming from our f_4; i.e.,[6]

$$T_{\mathrm{eff}} = \left[\frac{2^{n_c+2}}{3V_q}\pi^{(n_c-1)/2}\frac{\varphi(n_c+4)}{\varphi(4)}\Gamma\left(\frac{n_c+5}{2}\right)\right]^{1/4}A^{-n_c/4}T^{(n_c+4)/4}. \tag{16}$$

The corresponding effective four-dimensional entropy satisfies

$$(S_4R^3)_J \sim R^{3n_c/(3+n_c)}A^{-(3n_c/4)(2n_c+7)/(n_c+3)}. \tag{17}$$

This is the most important result of our paper. It implies that the whole process produces entropy in the four-dimensional world, which would be considerable if A decreases very rapidly, and/or R grows very fast, and/or the number of extra dimensions is quite large.

This approximation ($A \to \infty$, epoch I) is valid from the "big bang" on, up to a critical time, to be determined by the condition

$$AL = \beta_0 R^{3/(3+n_c)}A^{n_c/(3+n_c)}. \tag{18}$$

At epoch II only the $\vec{m}=0$ mode in the sum contributes, so that

$$S_D{}^{II} = \tfrac{1}{2}(n_c+2)S_4{}^{II}(LA)^{-n_c}. \tag{19}$$

This implies that from this critical time on, the expansion is adiabatic in the four-dimensional sense also,

$$(S_4R^3)_{II} = \frac{2L^{n_c}}{n_c+2}X, \tag{20}$$

so that all the effects of the extra dimensions on the four-dimensional physics have disappeared.

It is not difficult to find solutions with the desired general characteristics. Starting from Einstein's equations, which in our case reduce to

$$\begin{gathered}
3\left(\frac{\dot R}{R}\right)^2 + 3n_c\frac{\dot R}{R}\frac{\dot A}{A} + \frac{n_c}{2}(n_c-1)\left(\frac{\dot A}{A}\right)^2 = \lambda + 8\pi G\rho,\\
2\frac{\ddot R}{R} + \left(\frac{\dot R}{R}\right)^2 + 2n_c\frac{\dot R}{R}\frac{\dot A}{A} + n_c\frac{\ddot A}{A} + \frac{n_c}{2}(n_c-1)\left(\frac{\dot A}{A}\right)^2 = \lambda - 8\pi G p,\\
(n_c-1)\frac{\ddot A}{A} + \frac{(n_c-1)(n_c-2)}{2}\left(\frac{\dot A}{A}\right)^2 + 3(n_c-1)\frac{\dot R}{R}\frac{\dot A}{A} + 3\frac{\ddot R}{R} + 3\left(\frac{\dot R}{R}\right)^2 = \lambda - 8\pi G p,
\end{gathered} \tag{21}$$

it is very easy to check that when $n_c = 1$ a vacuum solution exists with zero cosmological constant, and such that $R \sim t^{1/2} \sim A^{-1}$ (this, however, is too slow a rate to be of any help with the horizon and flatness problems).

For both dust ($p=0$) and radiation in the large-n_c limit there is a solution of the type

$$\begin{gathered}
A \sim R^{-m},\\
R \sim [R_0{}^{mn_c/2} - (4\pi G\rho_0)^{1/2}t\,]^{-2/mn_c}.
\end{gathered} \tag{22}$$

The behavior $A \sim R^{-m}$ is not strictly compatible with the compactification of the extra dimensions, as this *Ansatz* is valid only during epoch I. On the other hand, in epoch II the extra dimensions essentially decouple, so that the universe is well described by the standard four-dimensional model. A really satisfactory solution should incorporate in an essential way the fact that C_{n_c} is a compact manifold, even during epoch I.

In an universe evolving in accordance with a scenario of this type the amount of entropy produced can be as large as $\Delta(S_4 R^3) = 10^{86}$ (which is the amount necessary to solve the horizon and flatness problems, see Guth[4]) and the Hubble constant H_0 can be fixed to be equal to the actual observed value of 50 km/sec Mpc $\sim 10^{-18}$ sec^{-1}. Of course this implies neglect of the duration of epoch II, which is very short in this model. (The universe is then much older than 10^{10} years!) It cannot be taken seriously as representing our universe though, because, for example, it predicts too large a variation with time of the fine-structure constant by many orders of magnitude (it is known observationally to be $\dot{\alpha} < 10^{-24}$ sec^{-1}, see Dyson[7]). Nevertheless, we think that its existence makes it plausible that a solution of Einstein's equations will exist with all the desired characteristics.

We want to stress, finally, that the fact that the effective coupling constants in four dimensions cannot have too large a variation with time is a major constraint in all Kaluza-Klein cosmologies. It implies that after a period of more or less quick contraction the extra dimensions must remain almost stationary—with a very small radius, presumably smaller than $10^{17} m_p^{-1}$ (see Ref. 1)—until now. It is amusing to remark, in this connection, that it is not possible for either vacuum, dust, or radiation, within the context of the given *Ansatz* (1) for the metric, to build a model in which the ordinary dimensions, independently of their number, expand and the extra ones remain strictly stationary. (It seems possible to have $\dot{A}$ small in our vicinity—for $z < 1000$, say—in such a way that $\dot{\alpha}$ is below the experimental limit; but dust or radiation models in which $\dot{A} = 0$ but $\dot{R} \neq 0$ must necessarily have a more complicated topology than the one assumed here.)

In conclusion, we have suggested a cosmological scenario for the production of large amounts of entropy in an universe undergoing dynamical compactification of its extra dimensions. The possibility of solving the horizon and flatness problems in this type of scenario is of course very exciting. Further research on these topics is currently under way.

We are especially grateful to Larry Abbott for his help and many fruitful discussions. This work was supported by the U. S. Department of Energy under Contract No. DE-AC02-76ER03230-A005 and by a Fulbright-MEC Grant.

---

[a] On leave of absence from Departamento de Física Téorica, Universidad Autónoma de Madrid, Madrid 34, Spain.

[b] On leave of absence from Laboratoire de Physique Théorique et Hautes Energies, University of Paris-Sud, 91405 Orsay, France.

[1] See, for example, A. Salam and J. Strahdee, Ann. Phys. (N.Y.) 141, 316 (1982).

[2] E. Cremmer and J. Scherk, Nucl. Phys. B103, 399 (1976); J. Scherk and J. H. Schwarz, Phys. Lett. 82B, 60 (1979).

[3] This idea has been suggested earlier by A. Chodos and S. Detweiler, Phys. Rev. D 21, 2167 (1980); P. G. Freund, Nucl. Phys. B209, 146 (1982).

[4] A. Guth, Phys. Rev. D 23, 347 (1981).

[5] We are adopting here the kinetic viewpoint that the dimensions of the momentum in $4+n_c$ dimensions are the same as in ordinary space-time, namely 1 in natural units. In other words, the "classical" definition $p^\mu \equiv m\, dx^\mu/ds$ is adopted.

[6] Other definitions are possible; one could, for example, define the effective temperature in such a way that the particle density is the same as the one stemming from f_4, i.e.,

$$T_{\text{eff}}' = \left[\frac{\pi^{(n_c+1)/2}}{4V_q} \frac{\varphi(n_c+3)}{\varphi(3)} \frac{\Gamma(n_c+3)}{\Gamma((n_c+3)/2)} \right]^{1/3} \times A^{-n_c/3} T^{(n_c+3)/3}.$$

If the original distribution function, f_4, is not very far from equilibrium all the possible effective temperatures, although different, should give roughly equivalent physical results. That this is the case for our f_4 can be easily checked by a direct comparison.

[7] F. J. Dyson, *Aspects of Quantum Theory*, edited by A. Salam and E. P. Wigner (Cambridge Univ. Press, Cambridge, England, 1972).

COSMOLOGY FROM HIGHER-DIMENSIONAL GRAVITY

Q. SHAFI
NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA
and International Centre for Theoretical Physics, 34100-Trieste, Italy

and

C. WETTERICH
CERN, Geneva, Switzerland

Received 13 June 1983

A scenario describing how the four-dimensional Friedmann cosmology may evolve from a model of gravity in $4 + D$ dimensions which admits spontaneous compactification of D dimensions is presented. In particular, the possibility that the transition to the standard hot big bang model involves an inflationary phase is explored.

1. Higher-dimensional cosmology. Kaluza and Klein [1] showed several decades ago how a unified description of gravity and electromagnetism becomes possible, starting from Einstein's theory of gravity in five dimensions. This idea has been generalized [2–4] to include also the weak and strong interactions. In these $(4 + D)$-dimensional models D spacelike dimensions are spontaneously compactified. The symmetries of this "internal space" appear as gauge symmetries of the effective four-dimensional theory. Internal spaces admitting the gauge group SU(3) × SU(2) × U(1) or grand unified groups have been proposed [3,4]. Even though a fully realistic model still remains to be found, the idea is elegant enough to warrant serious investigations of some of its implications.

In this letter we discuss the cosmology of models of higher-dimensional gravity. (For an earlier discussion, see ref. [5].) For time scales sufficiently large compared with the Planck time, the cosmology is well approximated by the standard four-dimensional Friedmann universe. An attempt is made to describe part of the transition from the early $(4 + D)$-dimensional universe to the standard cosmology which will be realized at later times. We will not address the problem of possible changes in the topology of space–time which may happen at an even earlier time during this transition. We rather investigate the time evolution of the characteristic length scales of the internal space and three-dimensional space, assuming that at this stage of the evolution space–time has the topology of the final Friedmann universe. In particular, we explore the possibility that the universe underwent an inflationary phase [6] during this transition. The inflation is characterized by an exponential expansion of three spacelike dimensions, while the characteristic length scale of the remaining D spacelike dimensions remains almost constant. Provided the exponential expansion lasts long enough, not only is the evolution of the asymmetry between the observed three spacelike dimensions and the internal dimensions described, but this transition also could provide a solution [6,7] to several major cosmological puzzles!

A small causally connected region could have expanded sufficiently to become the whole observed universe. This would explain the observed homogeneity of our universe and the remarkable isotropy of the 3 K background radiation. The curvature of the three space dimensions decreases exponentially during the inflationary phase, thus explaining why the universe today is so flat or, equivalently, has become so old (measured in Planck time units) and why the observed mass density is so near the critical density. Our preliminary results suggest that an inflationary scenario can indeed be implemented within the framework of higher-dimensional gravity.

Higher-dimensional gravity contains, after dimensional reduction, a theory of scalars coupled to gravity. During the early evolution of the universe, the characteristic length scale of the internal space will vary with time. This corresponds to a time varying scalar field which is a singlet with respect to the four-dimensional gauge group. Only when this scalar singlet attains its final vacuum expectation value (VEV) does the effective four-dimensional cosmological constant vanish. Prior to this, however, the value of the scalar field is different from its VEV, and an effective positive cosmological constant is present for the four-dimensional gravity theory. The observed three space dimensions therefore expand exponentially in a de Sitter phase. The central idea of the inflationary universe, proposed earlier in the context of GUT phase transitions, may now be realized in the present framework.

One possible scenario depicting how "primordial inflation" of three space dimensions can arise is described as follows: at very early times the characteristic length scales of the internal space (S) and of the three-dimensional space (R_3) are of the same order of magnitude. Both may have attained, during this early evolution, a value larger than the final vacuum value for S, $\langle S \rangle$. At some time, S reaches a maximum and starts to fall again. When S is near its maximum, it changes slowly with time. During this period, the effective four-dimensional cosmological constant is positive, since S is different from $\langle S \rangle$, and R_3 expands exponentially. When S is sufficiently far away from its maximum, it may again follow a fast evolution towards $\langle S \rangle$ and oscillate around $\langle S \rangle$. These oscillations are quickly damped through the production of matter and radiation, thereby heating the universe and making a transition to the usual four-dimensional Friedmann universe. The heating produces the large entropy observed in the present universe. The latter part of the scenario is in complete analogy to a scalar field falling down from a local extremum of its "de Sitter potential" (where contributions of curvature to the scalar potential are taken into account) into the minimum of the potential at zero four-dimensional curvature. One can indeed give a description in this language.

2. The model. In order to obtain a four-dimensional Friedmann universe at later times, the $(D + 4)$-dimensional model of gravity must fulfil several conditions [8]. For zero temperature, it must exhibit a "ground state" which is a direct product of four-dimensional Minkowski space and some internal space. The ground state solution must obey the $(D + 4)$-dimensional field equations. Excitations of the $(D+4)$-dimensional metric tensor above this ground state contain symmetric second rank tensors, vectors and scalars with respect to the four-dimensional general coordinate transformations. The four-dimensional gravitational interactions must be governed by a term linear in the four-dimensional curvature scalar with positive Planck's constant. The four-dimensional cosmological constant should vanish. The ground state must correspond at least to a local minimum of the scalar potential. Finally, all kinetic terms of the four-dimensional fields must have the right sign.

In pure $(D + 4)$-dimensional gravity theories, the above conditions can only be met [8] if terms containing more than two derivatives are included in the action[‡1]. As an example [8], we shall discuss the following action:

$$S = -\frac{1}{V_0}\int \mathrm{d}^d x\, \hat{e}(\alpha \hat{R}^2 + \beta \hat{R}_{\hat{\mu}\hat{\nu}} \hat{R}^{\hat{\mu}\hat{\nu}} + \gamma \hat{R}_{\hat{\mu}\hat{\nu}\hat{\sigma}\hat{\lambda}} \hat{R}^{\hat{\mu}\hat{\nu}\hat{\sigma}\hat{\lambda}} + \delta \hat{R} + \epsilon)\,, \tag{1}$$

$\hat{R}_{\hat{\mu}\hat{\nu}\hat{\sigma}\hat{\lambda}}$ is the d-dimensional curvature tensor and $\hat{e} = \sqrt{|g|}$ is the determinant of the vielbein. Coordinates are denoted by $x^{\hat{\mu}} = (x^{\mu}, y^{\alpha}) = (t, x^i, y^{\alpha})$ where x^i and y^{α} are spacelike coordinates of the three-dimensional space and the D-dimensional internal space. The normalization constant V_0 is chosen to be the volume of the internal space for $S = \langle S \rangle$. Then $\delta(\epsilon)$ has dimension mass^2 (mass^4) and α, β, γ are dimensionless.

The action (1) should be considered as an *effective* action at some distance scale close to the Planck scale. All contributions from quantum fluctuations which can be cast in the form (1) are supposed to be included (they may even be mainly responsible for the appearance of higher derivative terms). Of course, the full effective theory would also contain even higher derivative terms like R^4, etc. We should consider eq. (1) as an approximation which is valid so long as R is sufficiently smaller than the characteristic length scale $\delta^{-1/2}$. (It is difficult, at this stage, to give a precise estimate of the limits of validity, since this would require some knowledge about the coefficients of R^4 terms, etc., in the effective action.) Assuming that we can neglect these invariants, the remaining effects of the inclusion of quantum fluctuations will result in a scale dependence of the parameters $\alpha, \beta, \gamma, \delta$ and ϵ and in contributions violating general coordinate invariance in d dimensions due to spontaneous compactification. Both effects may be neglected here. The cosmology of this model is governed by the classical field equations derived from this action [9]:

[‡1] An alternative possibility is the introduction of supplementary matter fields in the $(D + 4)$-dimensional action. We expect that some aspects of the discussion will also apply to this kind of model.

$$\begin{aligned}
&2\alpha\hat{R}\hat{R}_{\hat{\mu}\hat{\nu}} + 2\beta\hat{R}_{\hat{\mu}\hat{\sigma}}\hat{R}_{\hat{\nu}}{}^{\hat{\sigma}} + 2\gamma\hat{R}_{\hat{\mu}\hat{\sigma}\hat{\tau}\hat{\rho}}\hat{R}_{\hat{\nu}}{}^{\hat{\sigma}\hat{\tau}\hat{\rho}} + \delta\hat{R}_{\hat{\mu}\hat{\nu}} \\
&\quad - \tfrac{1}{2}\hat{g}_{\hat{\mu}\hat{\nu}}(\alpha\hat{R}^2 + \beta\hat{R}_{\hat{\sigma}\hat{\tau}}\hat{R}^{\hat{\sigma}\hat{\tau}} + \gamma\hat{R}_{\hat{\sigma}\hat{\tau}\hat{\rho}\hat{\lambda}}\hat{R}^{\hat{\sigma}\hat{\tau}\hat{\rho}\hat{\lambda}} + \delta\hat{R} + \epsilon) \\
&\quad + 2\alpha(\hat{g}_{\hat{\mu}\hat{\nu}}\hat{R}_{;\hat{\sigma}}{}^{\hat{\sigma}} - \hat{R}_{;\hat{\mu}\hat{\nu}}) + \beta(\hat{g}_{\hat{\mu}\hat{\nu}}\hat{R}^{\hat{\sigma}\hat{\tau}}{}_{;\hat{\sigma}\hat{\tau}} + \hat{R}_{\hat{\mu}\hat{\nu};\hat{\sigma}}{}^{\hat{\sigma}} \\
&\quad - \hat{R}_{\hat{\mu}}{}^{\hat{\sigma}}{}_{;\hat{\nu}\hat{\sigma}} - \hat{R}_{\hat{\nu}}{}^{\hat{\sigma}}{}_{;\hat{\mu}\hat{\sigma}}) + 2\gamma(\hat{R}_{\hat{\mu}}{}^{\hat{\sigma}}{}_{\hat{\nu}}{}^{\hat{\tau}}{}_{;\hat{\sigma}\hat{\tau}} + \hat{R}_{\hat{\mu}}{}^{\hat{\sigma}}{}_{\hat{\nu}}{}^{\hat{\tau}}{}_{;\hat{\tau}\hat{\sigma}}) \\
&\quad = \tfrac{1}{2}\hat{T}_{\hat{\mu}\hat{\nu}} .
\end{aligned} \tag{2}$$

The quantity $\hat{T}_{\hat{\mu}\hat{\nu}}$ on the right-hand side of eq. (2) is the $(D+4)$-dimensional generalization of the energy–momentum tensor. It includes all incoherent excitations and vanishes for zero entropy.

The conditions stated previously for obtaining a four-dimensional Friedmann universe are met for the parameter values [8]

$$\begin{aligned}
&\varphi = D(D-1)\alpha + (D-1)\beta + 2\gamma > 0 , \\
&\beta + [4/(D-2)]\gamma > 0, \quad \gamma > 0, \quad \delta > 0 , \\
&\epsilon = \tfrac{1}{4}\delta^2 D(D-1)/\varphi , \quad (D-1)\beta + 2\gamma > 0 .
\end{aligned} \tag{3}$$

The field equations (2) admit the manifold $S^D \times \mathcal{M}^4$ as a zero temperature solution, where the radius $\langle S \rangle$ of the internal D-dimensional hypersphere S^D is given by

$$\langle S \rangle^2 = 2\varphi/\delta . \tag{4}$$

A four-dimensional effective theory is obtained by integrating out the coordinates of S^D. This effective theory exhibits $SO(D+1)$ gauge invariance. The ground state corresponds to an absolute minimum of the scalar potential. The Planck mass M_P is given by

$$M_P^2 = 16\pi[(D-1)\beta + 2\gamma]\delta/\varphi , \tag{5}$$

and the four-dimensional cosmological constant vanishes.

3. De Sitter solutions. For early times, R_3 and S are both of the order of the Planck length and the cosmology is best described by using directly the $(D+4)$-dimensional field equations (2). In general, this system of equations is quite complicated and admits a large variety of solutions. For simplicity, we will restrict ourselves to solutions where space is homogeneous and isotropic separately in the usual three dimensions and in the internal D dimensions. At a given time, space is a direct product of S^3 and S^D, but the radii $R_3(t)$ of S^3 and $S(t)$ of S^D are time dependent. [The discussion can easily be adapted to an open four-dimensional universe. $R_3(t)$ corresponds to the scale factor in the Robertson–Walker metric.] With this ansatz, one finds

$$\begin{aligned}
&\hat{R}_{\alpha\beta\gamma\delta} = -(S^{-2} + \dot{S}^2 S^{-2})(\hat{g}_{\alpha\gamma}\hat{g}_{\beta\delta} - \hat{g}_{\alpha\delta}\hat{g}_{\beta\gamma}) , \\
&\hat{R}_{ijkl} = -(R_3^{-2} + \dot{R}_3^2 R_3^{-2})(\hat{g}_{ik}\hat{g}_{jl} - \hat{g}_{il}\hat{g}_{jk}) , \\
&\hat{R}_{\alpha i\beta j} = -\dot{S}S^{-1}\dot{R}_3 R_3^{-1}\hat{g}_{\alpha\beta}\hat{g}_{ij} , \\
&\hat{R}_{\alpha 0\beta 0} = -\ddot{S}S^{-1}\hat{g}_{\alpha\beta} , \quad \hat{R}_{i0j0} = -\ddot{R}_3 R_3^{-1}\hat{g}_{ij} .
\end{aligned} \tag{6}$$

($\hat{R}_{\hat{\mu}\hat{\nu}\hat{\sigma}\hat{\lambda}} = 0$ for all other index combinations except those obtained by index permutations from the above expressions.) Inserting these expressions in the field equation (2), we obtain three independent equations [‡2] for $S(t)$ and $R_3(t)$.

Since we are especially interested in an exponential expansion of the usual three space dimensions, we first look for de Sitter solutions in four-dimensional space–time:

$$S(t) = S_H = \text{const.}, \quad \dot{R}_3^2/R_3^2 = H^2 - R_3^{-2} . \tag{7}$$

Indeed, such solutions do exist [‡3] (for $\hat{T}_{\hat{\mu}\hat{\nu}} = 0$) and the field equations (2) are fulfilled for

---

[‡2] In the following we will concentrate on solutions where temperature effects can be neglected ($\hat{T}_{\hat{\mu}\hat{\nu}} = 0$). For a more general discussion, the time dependent temperature is a third independent function and the three equations will have a still larger variety of solutions, depending on the initial conditions.

[‡3] De Sitter solutions in four-dimensional gravity theories including higher derivative invariants have been discussed by Starobinski [10].

$$H^2 = \tfrac{1}{24}\delta\,[D(D-1)/\varphi]\,(1-z)^2(1+\sigma z)^{-1}\,, \qquad (8)$$

where

$$z = \langle S\rangle^2/S_{\mathrm{H}}^2 \qquad (9)$$

is determined by

$$a_1 z^3 + a_2 z^2 + a_3 z + a_4 = 0\,, \qquad (10))$$

with

$$\begin{aligned}
a_1 &= D(D-1) + D(D-1)\rho + [D(D-1) - 12]\,\sigma \\
&\quad - 12\sigma^2\,, \\
a_2 &= -3D(D-1) - 3D(D-1)\rho \\
&\quad - (3/D)[D^2(D-1) - 4D + 16]\sigma \\
&\quad + 12[(D-4)/D]\sigma^2\,, \\
a_3 &= (3/D)\,[D^2(D-1) - 4D - 16] + 3D(D-1)\rho \\
&\quad + (3/D)\,[D^2(D-1) - 4D - 16]\,\sigma\,, \\
a_4 &= -D(D-1) + 12 - D(D-1)\rho \\
&\quad - [D(D-1) - 12]\sigma\,, \qquad (11)
\end{aligned}$$

and

$$\sigma = -D(D-1)\alpha/\varphi\,, \quad \rho = -(D-4)\beta/\varphi\,. \qquad (12)$$

The Hubble constant H is the characteristic scale for the exponential expansion. Compared with the Planck scale, it is given by

$$\begin{aligned}
H^2/M_{\mathrm{P}}^2 &= (1/384\pi)\{D(D-1)/[(D-1)\beta + 2\gamma]\} \\
&\quad \times (1-z)^2(1+\sigma z)^{-1}\,. \qquad (13)
\end{aligned}$$

Note that for not too small values of $(D-1)\beta + 2\gamma$ this ratio may be considerably smaller than 1. If z is also small, all relevant mass scales of such a de Sitter solution are well below the Planck mass! Note that for

$$\gamma = \tfrac{1}{2}(12/D - 3)\,[1 - 12/D(D-1)]^{-1}\beta\,, \qquad (14)$$

one has $a_4 = 0$ and therefore $z = 0$ is a solution of (10). We will come back to this interesting case when we consider values of γ near the value (14).

4. Primordial inflation. For a given set of parameters α, β, γ we may have up to three de Sitter solutions corresponding to the solutions of eq. (10). However, there is no special reason why the $(D+4)$-dimensional universe should have started in any of the de Sitter solutions. Imagine, however, that $S(t)$ and $R_3(t)$ are both roughly of order M_{P}^{-1} and evolve more or less randomly in phase space. Different regions of space–time may have different values of S and R_3. Assume now that at some time one space–time region comes close in phase space to one of the above de Sitter solutions. If this region stays long enough in the vicinity of such a solution the usual three space dimensions expand exponentially during this time and the whole observed universe could have evolved from such a region. The solution of the cosmological puzzles in the inflationary scenario requires that the time t_{c} during which the universe is in a phase of exponential expansion exceeds by a factor 60 or more the inverse Hubble constant (8):

$$t_{\mathrm{c}} \gtrsim 60H^{-1}\,. \qquad (15)$$

In order to estimate t_{c} we consider solutions of the field equations (2) which describe small deviations from the de Sitter solutions. We define

$$\begin{aligned}
s(t) &= \ln\,[S(t)/S_{\mathrm{H}}]\,, \\
v(t) &= \ln\,[R_3(t)/\bar{R}_3] - u(t)\,, \qquad (16)
\end{aligned}$$

where $u(t)$ corresponds to a de Sitter solution with $\dot{u}(t)^2 = H^2 - \bar{R}_3^{-2}\exp[-2u(t)]$. Since $u(t)$ grows with a rate given by H, we may neglect all factors $\exp[-u(t)]$ after a relatively short time in the de Sitter phase. Since we are interested in small deviations from the de Sitter solution, we first linearize the field equations in s and v. Furthermore, we are interested in solutions with a slow time dependence and we may therefore neglect the higher time derivatives appearing in the linearized field equations. Since $\dot{v} \sim Hs$, a consistent approximation to the field equations near a de Sitter solution is given as follows:

$$\begin{aligned}
&c_1 s + c_2\dot{s} + c_3\dot{v} + c_4\ddot{v} = 0 \qquad (\text{from } \hat{T}_{00} = 0)\,, \\
&c_1 s + (c_2 - b_2)\dot{s} + c_3\dot{v} + (c_4 + b_1)\ddot{v} = 0 \quad (\text{from } \hat{T}_{ij} = 0)\,, \\
&d_1 s + d_2\dot{s} + d_3\dot{v} + d_4\ddot{v} = 0 \quad (\text{from } \hat{T}_{\alpha\beta} = 0)\,. \qquad (17)
\end{aligned}$$

The constants b_i, c_i and d_i are calculated as functions of the parameters α, β, γ and δ.

This system of equations has a solution [‡4]

[‡4] For footnote see next page.

$$s = s_0 \exp(\omega t)\,, \quad v = (b_2/b_1) s_0 \exp(\omega t)\,, \tag{18}$$

which is within the validity of our approximations, provided $\omega \equiv t_c^{-1} \ll H$. In this case, it approximates the evolution towards or away from a de Sitter solution, depending on the sign of ω. A sufficient inflation requires $\omega/H \lesssim 1/60$. One finds

$$\omega/H = [4z(1+\sigma)/D]\,[(1+\sigma z)^2/(1-z)^2]\tilde{A}/\tilde{B}\,, \tag{19}$$

with

$$\tilde{A} = (1+\sigma z)(1-z) - (2/D)[1+2(1+\sigma)z+\sigma z^2]\,, \tag{20}$$

and $\tilde{B}$ a lengthy function of σ, and z typically of the order of unity. The critical quantity ω/H is small for small z, large D, small $\tilde{A}/\tilde{B}$ and σ near -1. Without fine tuning of parameters, it seems possible to obtain $\omega/H \lesssim 1/60$. Small values of z [see eq. (14)] are especially interesting, since in this case all length scales are large compared to the Planck length during the de Sitter phase.

To conclude, our results suggest that sufficient inflation can occur in theories of higher-dimensional gravity. More work is needed to establish a complete scenario of the transition to the Friedmann universe. The mechanism for the generation of entropy (heating the universe after the expansion) remains to be clarified. The scalar corresponding to S is a singlet of the gauge group and therefore decouples from the massless gauge bosons (and the chiral fermions in a more realistic model). However, in a more elaborate model some non-singlet scalars must have masses many orders of magnitude below the Planck mass in order to induce spontaneous symmetry breaking down to the $SU(3)_c \times U(1)_{em}$ gauge group. Provided the couplings of these scalars to the singlet are not too small, they will be copiously produced by the oscillations of S around $\langle S \rangle$. Their decay into gauge bosons and fermions will then quickly establish thermodynamic equilibrium. Other questions concern the generation of primordial density fluctuations and the possibility that a GUT phase transition may occur after the transition to the Friedmann universe.

Q.S. would like to thank Floyd Stecker and NASA/Goddard Space Flight Center for hospitality.

References

[1] T. Kaluza, Sitzungsber. Preus. Akad. Wiss. Berlin, Math. Phys. K1 (1921) 966;
O. Klein, Z. Phys. 37 (1926) 895.

[2] Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
J.F. Luciani, Nucl. Phys. B135 (1978) 111;
J. Scherk and J.H. Schwarz, Nucl. Phys. B153 (1979) 61;
A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316.

[3] E. Witten, Nucl. Phys. B186 (1981) 412.

[4] C. Wetterich, Phys. Lett. 110B (1982) 379.

[5] A. Chodos and S. Detweiler, Phys. Rev. D21 (1980) 2167.

[6] A. Guth, Phys. Rev. D23 (1981) 347.

[7] A.D. Linde, Phys. Lett. 108B (1982) 389;
A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. 48 (1982) 1220;
A. Albrecht, P.J. Steinhardt, M.S. Turner and F. Wilczek, Phys. Rev. Lett. 48 (1982) 1437.

[8] C. Wetterich, Phys. Lett. 113B (1982) 377.

[9] B.S. De Witt, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965).

[10] A. Starobinsky, Phys. Lett. 91B (1980) 99.

[‡4] Two linear combinations of the three equations (17) are fulfilled by the construction of the solution. The third will then give a restriction on the parameters α, β, γ, and δ for which the zero temperature approximation is valid for the solutions near the de Sitter solutions. Weakening this restriction, forces us to include temperature effects in our discussion. This makes the calculations somewhat longer, but we do not expect that our answer depends crucially on the zero temperature approximation.

TOWARDS A REALISTIC KALUZA–KLEIN COSMOLOGY ☆

Deshdeep SAHDEV
Department of Physics, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 7 November 1983
Revised manuscript received 9 December 1983

Einstein's equations are solved for a class of spatially homogeneous, anisotropic, perfect fluid, cosmological models in $1 + d + D$ dimensions (d, D arbitrary). It is found that they can in general lead to universes with 3 dimensions expanding and D contracting. Furthermore, assuming this process to be isentropic, the temperature T is found to eventually drop below $1/R$ (R = radius of the compact space). A scenario for how the universe can pass from the multi-dimensional phase to the usual Friedman-Robertson-Walker one at this point, is then outlined. The horizon problem for these models is examined. It is found that the entropy falling within a causal horizon at the point of dimensional reduction is essentially a free parameter.

The emergence of gauge invariance as a major guiding principle for the formulation of physical laws has made it appealing, not to say imperative, to seek a deeper basis for its presence. Some attempts have accordingly been made to understand gauge invariance as a long distance symmetry of a theory chaotic at short distances [1]. An alternative approach – that associated with the names of Kaluza and Klein [2][‡2] – has been to view it as a part of a wider symmetry. This line of attack elevates gauge invariance to the status of space–time invariance by postulating the existence of extra compact dimensions: Theories generally covariant with respect to the extended space–time and gauge invariant once compactification occurs.

If we take this interpretation seriously we find ourselves viewing our 4-dimensional space–time manifold as merely a cross-sectional subspace of the true world geometry. This statement is best understood in terms of a two-dimensional analog, viz. a space–time whose spatial surfaces of homogeneity consist of cylinders. If the cross section of these surfaces contracts to a microscopic radius R while the typical distance between points along the flat direction increases, we eventually get an expanding (1+1)-dimensional space–time, with a compact dimension at each point, which is unobservable to probes (wavelengths) large compared to R.

In like manner, following up on the Kaluza–Klein conjecture, we must necessarily see our universe today as having three expanding spatial dimensions and a number, D, of compact ones, which will open out to us only at very high energies, but which undoubtedly have played an active role in the early universe. If we find convincing evidence of the existence of these dimensions, the Kaluza–Klein route to the unification of gauge theories and gravitation becomes as physically necessary as it is aesthetically compelling. In fact, it becomes a matter of physical necessity to formulate *all* short distance theories in $1 + n$ (>3) dimensions.

In this letter we outline a cosmological scenario which produces a universe of the sort suggested above[‡2] We note that in this altered scenario the horizon problem is resolved, in that the entropy of a region falling within a causal horizon is essentially a free parameter.

We start from the observation that a universe with three spatially homogeneous, "large" expanding dimensions and D compact (quasi, if not totally) static

---

☆ Work was supported in part by the US Department of Energy under Contract No. EY-76-C-02-3071.

[‡1] For a modern version of the ideas, see ref. [3].

[‡2] Earlier work along these lines can be found in refs. [4,5].

ones is certainly very anisotropic between "large" and "small" dimensions. (For simplicity we shall assume all the compact dimensions to be spatial and will treat their number D as a parameter of the theory.) This anisotropy could have a dynamical origin, i.e. an isotropically expanding n-dimensional universe could at some temperature have dynamically selected out three dimensions for preferential expansion, but it is equally conceivable that the anisotropy persists as we evolve backwards in time all the way to a cosmological singularity.

In either case, we expect a period in the evolution of the early universe during which all dimensions are comparable and our spatial manifold is (at least locally) a direct product of two subspaces.

For this regime, we may assume a metric of the form [5][‡3]

$$g_{\mu\nu} = \begin{bmatrix} 1 & & \\ & -r^2 g_{mn} & \\ & & -R^2 g_{MN} \end{bmatrix} . \tag{1}$$

The g_{mn}- and g_{MN}-spaces are d and D $(d + D = n)$ dimensional respectively. We, furthermore, assume the universe to be radiation dominated at this stage and hence take the energy–momentum tensor to have a form appropriate to a perfect fluid in n dimensions viz.

$$T_{\mu\nu} = -p g_{\mu\nu} + (p + \rho) u_\mu u_\nu , \tag{2}$$

where $u_0 = 1, u_i = 0$. The pressure p and the density ρ satisfy the equation of state $\rho = np$.

The Einstein equations

$$R_{\mu\nu} = -8\pi \bar{G}\, T_{\mu\nu} - \{[g_{\mu\nu}/(n-1)]\, T_\lambda{}^\lambda\} \tag{3}$$

then become

$$d\ddot{r}/r + D\ddot{R}/R = -8\pi\bar{G}\rho , \tag{4a}$$

$$k_\mathrm{d}/r^2 + (\mathrm{d}/\mathrm{d}t)(\dot{r}/r) + (\mathrm{d}\,\dot{r}/r + D\dot{R}/R)(\dot{r}/r) = (8\pi\bar{G}/n)\rho , \tag{4b}$$

$$k_\mathrm{D}/R^2 + (\mathrm{d}/\mathrm{d}t)(\dot{R}/R) + (d\dot{r}/r + D\dot{R}/R)(\dot{R}/R) = (8\pi\bar{G}/n)\rho , \tag{4c}$$

with

$$\rho = \mathrm{const.}/S^{n+1} , \tag{5}$$

where $S^n = r^d R^D$ is the mean volume. This form for ρ implicitly assumes that the universe evolves adiabatically. $\bar{G}$ is the gravitational constant appropriate to n dimensions. It has dimensions of $(\mathrm{length})^{n-1}$ and is given by [7]

$$\bar{G} = G V_\mathrm{c} , \tag{6}$$

where V_c is the volume of the compact space today. We shall take this to be an $(n-3)$-sphere of radius R_c. Thus $V_\mathrm{c} = \{2\pi^{(n-1)/2}/\Gamma[(n-1)/2]\} R_\mathrm{c}^{n-3}$, where the actual value of R_c (or, equivalently, that of $M_\mathrm{KK} = R_\mathrm{c}^{-1}$) will be treated as a parameter of the theory. Finally, we shall set the curvature $k_d = 0$, for simplicity, and will scale R so that $k_\mathrm{D} = 1$.

Eqs. (4) can be solved by making the following set of substitutions:

$$(8\pi\bar{G}/n)\rho = (A^{n-1}/n^2) r_0^{d(n+1)/n} / S^{n+1}$$
$$= A^{n-1}/n^2 S(r_0 = 1)^{n+1} , \tag{7a}$$

$$S(x) = r_0^{d/n} A \sigma(x) , \tag{7b}$$

$$r(x) = r_0 A r(x) , \tag{7c}$$

$$\mathrm{d}t = nA\sigma(x)\,\mathrm{d}x = nS(r_0 = 1)\,\mathrm{d}x , \tag{7d}$$

where r_0 is an arbitrary scale factor,

$$A^{n-1} = [n^2/(n+1)]\, 8\pi F_n\, 2\pi^{(n-1)/2}/\Gamma[(n-1)/2]$$
$$\times (M_\mathrm{p}^2 M_\mathrm{KK}^{n-3})^{-1} [\mathcal{E}(r_0 = 1)/F_n]^{(n+1)/n} , \tag{8}$$

and, F_n is defined such that the entropy, $\mathcal{E}(r_0)$, is given by

$$\mathcal{E}(r_0) = F_n S^n(r_0) T^n . \tag{9}$$

We can now exchange eqs. (4) for the following system of equations expressed in terms of x:

$$\frac{r'}{r} = \frac{x + K}{\sigma^{n-1}(x)} , \tag{10a}$$

$$\sigma^{n-1}\sigma^{n-1\prime\prime} + [\mathrm{d}/D(n-1) - 1](\sigma^{n-1\prime})^2$$
$$- (2d/D)(x + K)\sigma^{n-1\prime} + (n-1)\sigma^{n-1}$$
$$+ [(n-1)d/D](x + K)^2 = 0 , \tag{10b}$$

$$[r(x)/\sigma(x)]^{2d/D} = [n(n-1)D]^{-1} [(n-1) - \sigma^{n-1\prime\prime}]/\sigma^{n-1} \tag{10c}$$

where K is a constant of integration.

This system of equations has two different sets of solutions, one for K = 0 and the other for $K \neq 0$.

[‡3] For a general treatment of spatially homogeneous models see ref. [6].

These sets are not linked by any obvious limiting procedure except for very special values of d and D[‡4].

We focus here on the $K = 0$ class[‡5]. In this case, $\sigma^{n-1}(x)$ has the following power series expanison at $x = 0$,

$$\sigma^{n-1}(x) = x^2\,[(n-1)/2 + a_1 x^2 + a_2 x^4 + a_3 x^6 + a_4 x^8 + O(x^{10})]\,, \tag{11}$$

where,

$$a_1 = [n(n-1)^2 D/24]\,[2/(n-1)]^{2d/D(n-1)}\,, \tag{12a}$$

$$a_2 = [4a_1^2/5(n-1)]\,[1 - 4d/D(n-1)]\,, \text{ etc.} \tag{12b}$$

This expansion provides a set of initial conditions for numerically integrating eqs. (10) away from $x = 0$. This numerical integration was carried out using a standard fourth-order Runge–Kutta routine, to yield solutions very typically of the form shown in fig. 1.

It is worth mentioning that for the special case $d = 1, D = 2, a_2 = 0$ and hence the power series (11) terminates at the x^6-term. In other words, for this case we have an exact solution [8]:

$$S = r_0^{1/3} A x (1 - x^2)^{1/2}\,, \tag{13a}$$

$$r = r_0 A x/(1 - x^2)^{1/2}\,, \tag{13b}$$

$$R = A x (1 - x^2)\,, \tag{13c}$$

$$t = A[1 - (1 - x^2)^{3/2}]\,, \tag{13d}$$

which turns out to capture the essential features of this entire class of solutions. These are as follows:

The Universe begins and ends in a singularity. At both singularities $S \to 0$ and $\rho \to \infty$. For every value of r_0, the typical distance, $r(x)$, between points in the flat space increases monotonically from 0 to ∞, while the radius, R, of the curves space increases from zero to a maximum and then drops back to zero.

As we take the whole process to be isentropic, the temperature of the radiation is given by

$$T = [\mathcal{E}(r_0 = 1)/F_n]^{1/n}\,[1/S(r_0 = 1)]\,. \tag{14}$$

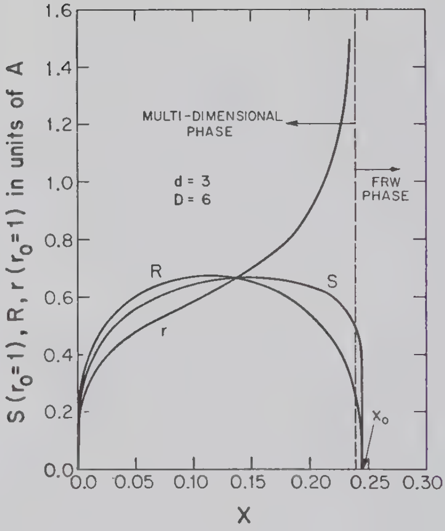


Fig. 1. The behavior of r, the scale factor for the usual dimensions, R, the radius of the compact ones and S, the mean radius, as functions of the modified time variable x, for $d = 3$, $D = 6$. The shapes are universal but the values and slopes at various points depend on d and D. The universe does not follow this solution out to the singularity at x_0. Rather, at some value of x, indicated by the dashed line, the temperature T falls below $1/R$ and the universe passes from the multi-dimensional phase to a FRW-one, in which it continues thereafter. The behavior of $T \propto 1/S$ is clear from the curve for S: the Universe cools from $T = \infty$, reheats until it changes phase and then cools once again, but as a FRW-universe. (The exact position of the dotted line depends on the parameters of the model.)

We assume that once the radiation temperature T drops below $1/R$ all excitations in the curved space decay to the flat one. The precise topologies for which this occurs are currently under investigation.

Once these excitations decay, the radiation is effectively confined to (i.e. propagates along) the d-dimensional space and the radiation pressure drops sharply along the curved directions. The compact space plays the role of spectator from here on. We, nevertheless, require it to stabilise in a period short enough not to cause any conflict with the current limits on the variation of G [10]. The details of how this stability is achieved are beyond the scope of the present discussion. We merely assume that whatever the stabilising mechanism (quantum gravity is a candidate), its net effect is to induce a constant effective pressure [‡6], $\lambda = 1/R_c^2$, for the compact space so that the equations of motion beyond the point, at which th extra dimensions "de-excite", read

---

[‡4] E.g. for $d = 1, D = 2, S^{n-1}(c = 0) = \lim_{c\to 0} S^{n-1}(c \neq 0)$. Cf. ref. [8]

[‡5] The qualitative features of the cosmological model discussed above are common to a much wider class of solutions and of $T_{\mu\nu}$'s. See ref. [9].

$$(\mathrm{d}/\mathrm{d}t)(\dot{r}/r) + (d\dot{r}/r + D\dot{R}/R)(\dot{r}/r) = 8\pi G(R_c/R)^{n-3} p^{(d)}, \tag{15a}$$

$$1/R^2 + (\mathrm{d}/\mathrm{d}t)(\dot{R}/R) + (d\dot{r}/R + D\dot{R}/R)(\dot{R}/R) = \lambda = 1/R_c^2, \tag{15b}$$

$$d\ddot{r}/r + D\ddot{R}/R = 8\pi G(R_c/R)^{n-3} \rho^{(d)}\ , \tag{15c}$$

where $p^{(d)}$ and $\rho^{(d)}$ refer to the d-dimensional pressure and density respectively ($\rho^{(d)} = \mathrm{d}p^{(d)}$).

Setting $R = R_c$, $\dot{R} = 0$, these equations clearly reduce to those for a d-dimensional Friedman–Robertson–Walker universe.

Because of the existence of a multi-dimensional phase prior to $T_{\text{de-ex}}$, the temperature profile of the early universe is radically altered. Had all dimensions been expanding, the radiation could not but cool. However, with some dimensions expanding and others contracting, it is the behaviour of the mean volume S^n which decides how the temperature changes. From fig. 1 it is clear that the universe starts out infinitely hot, cools to a certain $T_{\min}$ and then reheats until it becomes effectively (1 + 3)-dimensional. Beyond this point, the radiation temperature drops in the usual FRW-manner.

The horizon size at a given temperature, the lifetime of the multi-dimensional phase, $T_{\min}$, $T_{\text{de-ex}}$ and all other quantities relating to these models are determined by $\mathcal{E}(r_0 = 1)$ and M_{KK}, which are the only free parameters for these cosmologies.

To get reasonable estimates for these quantities, we need, in addition to the expansion (11) around $x = 0$, and the numerical solution, a power series for S^{n-1} near $x = x_0$, the second singularity:

$$S^{n-1}(y) = r_0 d(n-1)/n\, A^{n-1} \times [c_1 y + \tfrac{1}{2}(n-1)y^2 + c_2 y^m + \mathrm{O}(y^{m+1})]\ , \tag{16}$$

where

$$y = x_0 - x\ , \tag{17a}$$

$$c_1 = x_0\{d/D + [d(n-1)/D]^{1/2}\}/[1 - d/D(n-1)] = x_0/\beta\ , \tag{17b}$$

$$m = 3 - 2d/D(n-1) - 2d\beta/D\ , \tag{17c}$$

and c_2 has to be fit to the numerical solution.

The assumption of isotropy of pressure is now seen to break down only near $x = x_0$: $R/S = RT \times \mathcal{E}(r_0 = 1)^{-1/n}$ starts at a value of $[(n-1)/2]^{d/D(n-1)} \sim \mathrm{O}(1)$ at $x = 0$ and decreases monotonically to zero at $x = x_0$. The point $RT = 1$, which also gives to the "de-excitation" temperature, corresponds to $y^{-(3-m)/2} \simeq [\mathcal{E}(r_0 = 1)/F_n]^{1/n}$ and hence

$$T_{\text{deex}} = M[\mathcal{E}(r_0 = 1)/F_n]^{2[1/(3-m)-1]/n(n-1)}\ , \tag{18}$$

where $M = (M_{\mathrm{p}}^2 M_{\mathrm{KK}}^{n-3})^{1/(n-1)}$ [$(3-m)^{-1} = 0.82$, 1.80, 2.46 for $d = 3$, $D = 2, 4, 6$ respectively].

In contrast, the assumption of thermal equilibrium which requires that the reaction rate, $\sim\alpha^2 T$, exceed the expansion rate, $\dot{S}^n/S^n$, breaks down near both singularities. From the power series for S^{n-1}, we infer that equilibrium is lost for x^{-1} and y^{-1} greater than

$$x_l^{-1},\ [(n-1)y_l]^{-1} \simeq \alpha^2[\mathcal{E}(r_0 = 1)/F_n]^{1/n}\ , \tag{19}$$

where α is the coupling constant of the interaction responsible for equilibrium over the relevant temperature range. (We shall take $\alpha = 1/40$.)

Furthermore, the lifetime of the multi-dimensional phase

$$\tau = An \int_0^{T_{\max}} \sigma(x)\,\mathrm{d}x \sim A$$

while

$$T_{\min} \sim M[\mathcal{E}(r_0 = 1)/F_n]^{-2/n(n-1)}\ .$$

Finally, the horizon, $l(x)$, along the non-compact dimensions is given by

$$\frac{l(x)}{r(x)} = \int_0^x \frac{\mathrm{d}t}{r(t)} = \frac{n}{r_0} \int_0^x \frac{n(n-1)D\sigma^{n-1}(x)\,\mathrm{d}x}{[(n-1) - \sigma^{n-1\prime\prime}(x)]}\ , \tag{20}$$

where the integral was numerically found to exceed 1 well before $x = x_0 - y_l$. It should be noted that $l(x)/r(x)$ is independent of $\mathcal{E}(r_0 = 1)$ in sharp contrast to $l(t)/r(t)$ for the FRW-case, where [11].

$$l(t)/r(t) = r_0^{-1}(2\pi F_3)^{-1/2}[F_3/\mathcal{E}(r_0 = 1)]^{1/3} M_{\mathrm{p}}/T\ .$$

For the multi-dimensional scenario, the amount of entropy which falls into a causal horizon at T_{deex} can be chosen at will.

For $\mathcal{E}(r_0 = 1) \sim 10^{83} F_n$, the entropy of the observed universe [‡7], and $n = 9$, e.g., $T_{\text{deex}} = M \times 10^{3.4}$. If we require that $T_{\text{deex}} < M_{\text{p}}$ then we must choose $M_{\text{KK}} \lesssim 10^{14}$ GeV. For these values of M_{KK} and $\mathcal{E}(r_0 = 1)$, $x_l \simeq 10^{-6}$, $y_l \simeq 10^{-7}$, $y_{\text{deex}} \simeq 10^{-45}$, $T_{\min} = 10^{13}$ GeV and $\tau \simeq 10^{-3}$–10^{-4} GeV^{-1} [‡8].

[‡6] For details of the horizon problem, see ref. [11].

Thus the assumptions of the model hold well into the region of contraction for the extra dimensions. Even though the quantitative statements in the regime close to the singularities of the multi-dimensional phase cannot be taken too seriously (as for any cosmology), the qualitative features should survive all refinements.

It is hoped that this model, together with earlier work done along these lines [4,5], not only renders the concept of extra dimensions plausible, but also provides a basis for studying cosmological questions taking the extra dimensions into account.

Helpful conversations with H.A. Weldon, C. Ftaclas, B. Sathiapalan, S. Ramaswamy, Paul Steinhardt and S. Bludman are grateful acknowledged.

References

[1] D. Foerster, H.B. Nielsen and M. Ninomiya, Phys. Lett. 94B (1980) 135.

[2] T. Kaluza, Sitzungsber. Preuss Akad. Wiss., Phys. Math. K1 (1921) 966;
O. Klein, Z. Phys. 37 (1926) 875; Nature (London) 118 (1926) 516.

[3] E. Cremmer and J. Scherk, Nucl. Phys. B103 (1976) 399;
E. Witten, Nucl. Phys. B186 (1981) 412;
A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316, and references therein.

[4] A. Chodos and A. Detweiler, Phys. Rev. D21 (1980) 2167;
P.G.O. Freund and M.A. Rubin, Phys. Lett. 37B (1980) 233;
T. Dereli and R.W. Tucker, Phys. Lett. 125B (1983) 133.

[5] P.G.O. Freund, Nucl. Phys. B209 (1982) 146.

[6] M.P. Ryan and L.C. Shepley, Homogeneous relativistic cosmologies (Princeton, U.P., Princeton, 1975).

[7] S. Weinberg, Phys. Lett. 125B (1983) 265.

[8] R. Kantowski, Ph.D. Thesis (University of Texas);
R. Kantowski and R.K. Sachs, J. Math. Phys. 7 (1966) 443.

[9] D. Sahdev, in preparation.

[10] L. Halpern, ed., On the measurement of cosmological variations of the gravitational constant (University Presses of Florida, Gainesville, 1978).

[11] A. Guth, Phys. Rev. D23 (1981) 347.

[12] E. Alvarez and M. Belén Gavela, Phys. Rev. Lett. 51 (1983) 931.

[‡7] It is, of course, unaesthetic to introduce such a large number for $\mathcal{E}(r_0 = 1)$. However, it is worth pointing out that the parameter to really compare with the natural mass scale M irt these models is $A^{-1} \sim M[\mathcal{E}(r_0 = 1)/F_n]^{-(n+1)/n(n-1)}$. For $\mathcal{E}(r_0 = 1) = 10^{83} F_n$, $A^{-1}/M = 10^{-54}$ for $n = 3$ and 10^{-11} for $n = 9$. Thus on reduction to (1+3)-dimensions the relative effect of $\mathcal{E}(r_0 = 1)$ is vastly amplified. For a somewhat diffeent discussion of the entropy problem in such models, see ref. [12].

[‡8] These numbers are specific to $n = 9$. For $n = 15$, e.g., $T_{\text{de-ex}} = M \times 10^{2.2}$ so that here we can with impunity take M_{KK} to be as high as 10^{16} GeV, which is closer to the value suggested by ref. [7]. Note, however, that there is really no compelling reason for "deexcitation" to occur before the Universe reheats to T_{Planck}; only that if it does, the classical description given above would remain essentially valid over the entire lifetime of the multi-dimensional phase. At our present level of understanding of exact topology etc. of the compact dimensions it is clearly prudent to keep an open mind on the value of M_{KK}.

DIMENSIONAL REDUCTION IN THE EARLY UNIVERSE: WHERE HAVE THE MASSIVE PARTICLES GONE?

Edward W. KOLB
Theoretical Astrophysics Group, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA

and

Richard SLANSKY
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 1 November 1983

In many models based on reduction from greater than four dimensions, there are absolutely stable particles with masses of order R^{-1} (R is the compactification scale). If the temperature of the universe were ever close to R^{-1}, these massive states would have been present and some would have survived annihilation. We calculate the present mass density due to these particles and find both the five-dimensional model and some versions of $N = 8$ supergravity to be unacceptable. We discuss some possible solutions to this problem.

One of the most attractive approaches for unifying gauge theories with gravitation is based on enlarging the dimensionality of space–time. This program is traced back to the work of Kaluza [1] and Klein [2], who porposed that gravitation and electrodynamics in four dimensions might be unified as a pure gravitational theory in five dimensions. The vacuum geometry determines the effective low-energy theory; five-dimensional gravity reduces to four-dimensional gravity and a peculiar version of electrodynamics if the vacuum is the space $M^4 \times S^1$ (M^4 is four-dimensional Minkowski space and S^1 is the 1-sphere or circle).

Recent work has centered on constructing higher-dimensional theories that include non-abelian gauge interactions [3]. For example, it is possible to formulate $N = 8$ supergravity as a Kaluza–Klein theory in 11 dimensions [4] [‡1]. Another example is the quantum superstring, which must be formulated in 10 dimensions [6]. Although there has not yet emerged a low-energy theory with the observed gauge and fermionic structure of the standard electroweak-QCD model, the attractiveness of these approaches has generated much interest.

Typically, the vacuum geometry of interest is a direct product of M^4 with a compact space that has a high degree of symmetry. Then each field has a harmonic expansion about the vacuum into four-dimensional fields times the mass eigenfunctions of the extra dimensions. The zero modes correspond to the low-mass particle spectrum; the infinite sequences of higher modes have masses of order R^{-1}. It is usually assumed that R is of order $G_{\mathrm{N}}^{1/2} = m_{\mathrm{Pl}}^{-1} = 1.62 \times 10^{-33}$ cm times some power of the gauge coupling (G_{N} is Newton's constant), although there are models where R is determined by the electroweak breaking scale $G_{\mathrm{F}}^{1/2} = 6.74 \times 10^{-17}$ cm (G_{F} is Fermi's constant) [7].

We discuss a possible cosmological consequence resulting from the "four-dimensional particles corresponding to the non-zero modes of the harmonic expansions in mass eigenstates of the higher-dimensional fields", generically called "pyrgons" [‡2]. If the universe were ever at a temperature comparable to R^{-1}, the pyrgons would have been present. In the five-dimensional theory with an unbroken local U_1 symmetry, the pyrgons cannot decay solely into zero modes, although they can annihilate with antipyrgons. The stable pyrgons that survive annihilation contribute to the present energy density of the universe. We calculate the present number density of the pyrgons, and then require their contribution be less than the observed bounds.

[‡1] For a review of supergravity, see ref. [5].

[‡2] From the Greek πυργος, for tower or ladder; we acknowledge the grammatical license.

If annihilation were negligible, then today the number density of pyrgons would be comparable to the photon number density, which is acceptable only if the pyrgon mass satisfies $m_\psi \lesssim 100$ eV, just as for neutrino masses [8]. Thus, the pyrgons must be annihilated; we require that the annihilation rate ($\Gamma_A = n_\psi \sigma_A |v|$, where n_ψ is the pyrgon number density and σ_A is the annihilation cross section) must be comparable to the expansion rate of the universe ($\Gamma_E = T^2/m_{P1}$, where T is the temperature). If we assume $\sigma_A |v| = \alpha^2/m_\psi^2$, then at $T = m_\psi$ and $n_\psi = T^3$, we estimate $\Gamma_A/\Gamma_E = \alpha^2 m_{Pl}/m_\psi$. In the five-dimensional Kaluza–Klein model, $\alpha = R_{Pl}^2/R^2 = m_\psi^2/m_{Pl}^2$, and $\Gamma_A/\Gamma_E = m_\psi^3/m_{Pl}^3$. For $m_\psi < m_{Pl}$, the ratio is less than unity, and annihilation is ineffective at ridding the universe of stable pyrgons. However, if we modify the theory so the pyrgons carry a similar but larger charge such that α appearing in σ_A is order unity for any m_ψ, then $\Gamma_A/\Gamma_E \sim m_{Pl}/m_\psi$. If $m_\psi \ll m_{Pl}$, annihilation can be effective enough to reduce the pyrgon density below present observational bounds. For this to occur we will show $m_\psi \lesssim 10^6$ GeV.

The presence of absolutely stable pyrgons may appear to be a special feature of the five-dimensional model, resulting from the fact that the zero modes do not carry the U_1 charge. However, there are somewhat more realistic theories with stable pyrgons in which the zero modes do carry the gauge quantum numbers: for example, eleven-dimensional supergravity with vacuum $M^4 \times S^7$ has stable pyrgons even though most of the 256 zero modes do carry the SO_8 quantum numbers of the symmetry of S^7, as we show later.

We now turn to the details of the five-dimensional model, where the vacuum is $M^4 \times S^1$, and the harmonic expansion is just a Fourier series in the extra coordinate ($y = 2\pi R\theta$) with coefficients that are fields on M^4. (Generalizations will be indicated as we go along.) Thus, the fields are expanded as

$$\Psi_i(x,y) = \sum_{k=-\infty}^{\infty} e^{ik\theta} \phi_i^k(x), \tag{1}$$

where i is a space–time index and $|k|$ labels the mass eigenstate. (For a compact manifold with symmetry group G, the $e^{ik\theta}$ are replaced by representation matrices of G, $D^k(L_y^{-1})$, where L_y is the element of G that parameterizes the point y of the manifold. The sum on k is replaced by a sum or restricted sum over the representations of G.)

The equation of motion for small disturbances about the ground state geometry is given by [9]

$$\Box^{(5)} \Psi_i(x,y) = 0, \tag{2}$$

where $\Box^{(5)}$ is the five-dimensional d'alembertian. In the free field limit, each field $\phi_i^k(x)$ satisfies a wave equation,

$$(\Box^{(4)} + M_k^2)\phi_i^k(x) = 0, \tag{3}$$

where, in this case, the mass squared operator is just the d'alembertian in the extra dimension. Therefore, each term in the harmonic expansion (1) corresponds to a particle with mass $M_k^2 = (k/2\pi R)^2$. The generator of the charge is just $i\partial_y$, so the charge of $\phi_i^k(x)$ is proportional to k. The mass spectrum in four dimensions [9] contains a massless, neutral spin-2 particle; a massless, neutral spin-1 particle; a massless neutral scalar; and an infinite tower of charged spin-2 pyrgons with masses $M_k^2 = (k/2\pi R)^2$, $k = 1, 2, \ldots$.

Consider the decay of the pyrgons. We label each four-dimensional field in the harmonic expansion by the quantum number k. The amplitude for the process,

$$\phi^k \to \phi^{k_1} + \phi^{k_2} + \ldots + \phi^{k_n}, \tag{4}$$

is contained in a term of the five-dimensional effective action of the form,

$$I \sim 2\pi R \int d^4x \int d\theta \exp[i(-k + k_1 + \ldots + k_n)\theta] \times \phi^k(x)^* \phi^{k_1}(x) \ldots \phi^{k_n}(x), \tag{5}$$

where the contraction on space–time indices is implicit in the notation. Upon integration over the extra dimension, the decay rate for (4) is proportional to the Kronecker δ, $\delta(k_1 + k_2 + \ldots + k_n - k)$. The appearance of derivative couplings in the action still leads to the same δ-function, since the derivative does not mix modes. The existence of the δ-function means that no pyrgon ($|k| > 1$) can decay to zero modes ($k = 0$) only. [For spaces with higher symmetry, the generalization of the Kronecker δ is a $(3n - 3)$-symbol, which is nonzero only if it satisfies certain "triangle inequalities".] Of course, the annihilation of a pyrgon with its antiparticle can yield all zero modes.

We now calculate the number density n_ψ of remnant pyrgons from the big bang. There is no need to follow the evolution of the universe up to the time of compactification t_c if the initial conditions for n_ψ can be set at a time near t_c. At t_c the universe has become approximately four-dimensional and the excitations of the vacuum geometry may be reinterpreted as four-dimensional zero modes and pyrgons. If the four-dimensional temperature T at t_c is near enough to $T_c \equiv R^{-1}$, then the calculation of n_ψ is insensitive to the initial conditions, as will be seen. Otherwise, the calculation places a *lower* limit on n_ψ.

Unless very special initial conditions control the evolution of the universe for $t < t_c$, there will be pyrgons at t_c, and the initial condition for the ratio, $r = n_\psi/n_\gamma$, will be of order unity, since, typically, the excitations of the five-dimensional fields will be distributed over the modes in (1). Moreover, if the excited pyrgons ($|k| > 1$) decay rapidly enough into stable $|k| = 1$ pyrgons and zero modes, they will be present in a thermal distribution. The stable pyrgons decrease in number only by annihilation. For simplicity we ignore the $|k| > 1$ populations and compute n_ψ for the stable ones, assuming n_ψ for $k = +1$ and $k = -1$ are equal. Then n_ψ satisfies the equation,

$$\dot{n}_\psi = [(n_\psi^{\mathrm{eq}})^2 - n_\psi^2]\,\sigma_A|v| - \Gamma_E n_\psi\,, \tag{6}$$

where the equilibrium number density n_ψ^{eq} is determined by m_ψ and T. As is typical in gauge theories, for $T < m_\psi$, the scale of the pyrgon–antipyrgon annihilation cross section is set by m_ψ^{-2}:

$$\sigma_A|v| \simeq (\alpha^2/m_\psi^2)|v| \simeq \alpha^2 R^2/\sqrt{RT}\,, \tag{7}$$

where $m_\psi \simeq R^{-1}$. The four-dimensional expansion rate of the universe Γ_E is $(\rho_\psi + \rho_\gamma)^{1/2}/m_{\mathrm{Pl}}$, where ρ_ψ is the ψ energy density ($\rho_\psi = m_\psi n_\psi$ for $T < m_\psi$) and ρ_γ is the radiation energy density ($\rho_\gamma \sim T^4$). We calculate the ratio r from n_ψ in (6) and the photon number density $n_\gamma = T^3$ [‡3].

It is typical of all Kaluza–Klein theories, including the five-dimensional case, that the charge carried by the pyrgons is $\alpha = R_{\mathrm{Pl}}^2/R^2$, so $\sigma_A|v| = R_{\mathrm{Pl}}^4/R^2\sqrt{RT}$. Because of the small annihilation cross section, the final value of $r = n_\psi/n_\gamma$ will be near unity, independent of R. Thus, the model predicts primordial pyrgons to be nearly as abundant as primordial photons.

Although it is not in the spirit of Kaluza–Klein theories, one may make the *ad hoc* assumption that the pyrgons carry an additional charge that is similar to α, but not strongly dependent on R_{Pl}/R. If we assume for this charge that α is constant, then $\sigma_A|v| = \alpha^2R^2/\sqrt{RT}$, and the annihilation rate *increases* with R.

The ratio $r = n_\psi/n_\gamma$ is shown as a function of T/T_c in fig. 1 [‡4]; α is set to unity and the calculation is done for various values of R with the initial condition, $r = 1$. In the very special case that $r = 0$ at t_c, r rises quickly to the envelope of the curves; if $r > 1$ initially, this calculation gives a lower limit on the remnant pyrgons. After decoupling, r is constant in an isentropic expansion.

The calculation could be improved technically by considering pyrgon–antipyrgon capture into Coulomb bound states with large principle quantum numbers and subsequent annihilation [11,13], or by considering three-body initial states [14]. However, on the basis of the effect on monopole–antimonopole annihilation, we do not expect the improvements to change substantially the results in fig. 1.

Today ($T_\gamma \sim 10^{-13}$ GeV, $n_\gamma \sim 400\ \mathrm{cm}^{-3}$) the energy density contributed by the photons is $\rho_\gamma \sim n_\gamma T_\gamma \sim 10^{-4}\,\rho_c$, where ρ_c is the closure density, $10^{-29}\ \mathrm{g\ cm}^{-3}$. Since the total energy density of the universe is less than $2\rho_c$ [15], the present energy density of the ψ, $\rho_\psi = m_\psi n_\psi = R^{-1} n_\psi$, must satisfy $\rho_\psi/\rho_\gamma \lesssim 10^4$, or

$$r = n_\psi/n_\gamma \lesssim 10^4\, T_\gamma R\,. \tag{8}$$

Using the results from fig. 1, this limit is satisfied only if $R > (10^6\ \mathrm{GeV})^{-1}$ at $r \lesssim 10^{-14}$. If the annihilation cross section is changed, for instance, by using

---

[‡3] We will make the assumption that there is only one degree of freedom in the massless particle species and that the entropy produced in the annihilation is small.

[‡4] The solution to (6) has been studied for nucleons in ref. [10], for magnetic monopoles in ref. [11] and given in general in ref. [12].

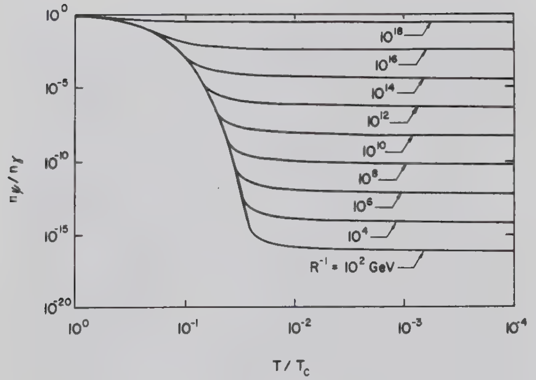


Fig. 1. The ratio r of the number density of the pyrgons n_ψ to the photon number density n_γ as a function of T/T_c for various values of the compactification scale R; $\alpha = 1$.

a different value of α by including more annihilation channels, the bound scales as $R > (\sigma_A m_\psi^2 \, 10^6 \text{ GeV})^{-1}$.

In summary, the above radius is much larger than the Planck radius. If the radius were R_{Pl} and the expansion of the universe were isentropic after ψ decoupling, the photon density would equal the density of charged spin-2 pyrgons with masses of order m_{Pl}. Moreover, this catastrophic prediction follows from other models with stable pyrgons.

Two possibilities for circumventing this bound on R come to mind. The first way is to relax the assumption that n_γ is constant. If a large amount of entropy were created after compactification, it would be possible to dilute the value of r to an acceptable level. However, the baryon asymmetry would be diluted by the same amount, so it is reasonable to require the entropy generation at an epoch prior to baryon number generation. This would seem to require compactification at energy scales $>10^{14}$ GeV. Another possibility is to relax the assumption that R is constant during the entire evolution of the universe. Indeed, Chodos and Detweiler [16] have discussed an interesting cosmological solution of the five-dimensional theory in which the extra dimension is "large" at early times, and subsequently shrinks as the other three spatial dimensions grow. If today $m_\psi = m_{\text{Pl}}$, then n_ψ / n_γ must be less than 10^{-28}, which requires $R \gtrsim (10^{-10} \text{ GeV})^{-1}$ at decoupling.

Since the five-dimensional model is very schematic, we now ask if these results can be generalized to more realistic theories with extra dimensions. The answer to the crucial question "are there stable pyrgons?" is model dependent. There are many types of higher dimensional theories: various versions of supergravity in 11 dimensions; superstrings in 10 dimensions [‡5]; pure Einstein gravity in any number of dimensions, where one might use a non-standard action [17] or a non-standard ansatz for the vacuum [18]. In some models there are "external" matter fields present to force the compactification, while in others, the compactification is due to the presence of a cosmological constant.

In eleven-dimensional supergravity with vacuum $M^4 \times S^7$, the zero modes and pyrgons are classified by helicity and SO_8. We show that if the SO_8 is unbroken or broken in a specific way, there are stable pyrgons. This can be seen as follows: the representations of SO_8 fall into 4 nonoverlapping classes that are congruent to the $\mathbf{1}, \mathbf{8}_v, \mathbf{8}_s$, or $\mathbf{8}_c$ [‡6]. All zero modes are in the $\mathbf{1}$ and $\mathbf{8}_s$ congruency classes [20,21]. The zero modes of helicity 2, 3/2, 1, 1/2, and 0, respectively, are in the $\mathbf{1}, \mathbf{8}_s, \mathbf{28}, \mathbf{56}_s$, and $\mathbf{35}_v + \mathbf{35}_c$, where $\mathbf{1}, \mathbf{28}$, and all three $\mathbf{35}$'s are in the $\mathbf{1}$ class, and $\mathbf{8}_s$ and $\mathbf{56}_s$ are in the $\mathbf{8}_s$ class. Any tensor product of any number of representations in the $\mathbf{1}$ and $\mathbf{8}_s$ classes remains in the $\mathbf{1}$ and $\mathbf{8}_s$ classes; it is impossible to reach representations in the $\mathbf{8}_c$ or $\mathbf{8}_v$ classes in this way. Thus, by the generalization of (5), if there are any pyrgons in the $\mathbf{8}_c$ or $\mathbf{8}_v$ classes, then some must be stable.

The pyrgon spectrum for S^7 is easily computed. The $\mathbf{1}$ graviton induces a harmonic expansion of mass eigenstates (1), with the sum on k in (1) over all SO_8 representations containing an SO_7 singlet: $\mathbf{1}, \mathbf{8}_v, \mathbf{35}_v, \mathbf{112}_v, \ldots$, or in terms of Dynkin labels, (0000), (1000), (2000), (3000), ... The $(2k+1,0,0,0)$ representations are all in the $\mathbf{8}_v$ class, but the $(2k,0,0,0)$ representations are in the $\mathbf{1}$ congruency class. (See ref. [18] for a review.) The supermultiplet accompanying the $(k000)$ term in the harmonic expansion for the graviton is a typical representation, which means that it is obtained by multiplying the zero mode representation by $(k000)$. Now, the product of an $\mathbf{8}_v$-type representation with a representation in the $\mathbf{1}, \mathbf{8}_v, \mathbf{8}_s$, or $\mathbf{8}_c$ class, respectively, is in the $\mathbf{8}_v, \mathbf{1}, \mathbf{8}_c$, or $\mathbf{8}_s$ class. Thus, the pyrgons in the supermultiplets where the spin-2 member transforms as $(2k + 1, 0, 0, 0)$ are in the $\mathbf{8}_v$ and $\mathbf{8}_c$ classes. For example, the "first" excited modes with helicity 2, 3/2, 1, 1/2, 0, respectively, are $\mathbf{8}_v, \mathbf{8}_c + \mathbf{56}_c, \mathbf{8}_v + \mathbf{56}_v + \mathbf{160}_v, \mathbf{8}_c + \mathbf{56}_c + \mathbf{160}_c + \mathbf{224}_{vc}, \mathbf{8}_v + \mathbf{56}_v + \mathbf{160}_v + \mathbf{112}_v + \mathbf{224}_{cv}$.

[‡5] Massive stable string configurations are expected in some versions of type 11 superstrings. We thank John Schwarz for this observation.

[‡6] We follow the SO_8 group theory conventions of ref. [19], except the subscripts "s" and "v" are interchanged, as in ref. [21].

If the SO_8 is broken, but broken without collapsing the congruency-class distinctions, then there still remain stable pyrgons. The most likely breaking patterns correspond to the little groups of the spinless zero modes, $\mathbf{35}_v + \mathbf{35}_c$ [18]. There are 4 little groups of the **35**'s: $SU_4 \times U_1$ and $SU_2 \times SU_2 \times SU_2 \times SU_2$, which preserve all the congruency class distinctions; and SO_7 and $SO_3 \times SO_5$, which collapse the distinctions. (Explicitly, the $\mathbf{8}_v$ and **1** classes are collapsed to a **1** class, and the $\mathbf{8}_s$ and $\mathbf{8}_c$ classes are collapsed to a spinor class.) Thus, in the former case, some pyrgons remain stable, but in the latter case, which includes the "squashed seven-sphere" with symmetry $SO_3 \times SO_5$, we conclude that all pyrgons are probably unstable [‡7].

In conclusion, interesting higher-dimensional theories may have stable pyrgons with masses of order R^{-1}; their cosmological implications can provide an important constraint on model building. The crucial ingredients for computing their contribution to the energy density of the universe are the structure of the harmonic expansion, the identification of the zero models and conservation laws, and the R dependence of σ_A. If there are stable pyrgons, then they become (yet further) candicates to dominate the dark matter of the universe.

[‡7] There are many examples where the stability of the pyrgons can be proven by congruency class arguments. For example, if all zero modes satisfy the usual electric charge–color relation, but some pyrgon does not.

This work was supported in part by the Department of Energy at Los Alamos and Fermilab, and by NASA at Fermilab. We are grateful to J. Harvey, W. Bardeen, C. Zachos, N. Metropolis and participants at the 1983 Aspen Center for Physics for useful discussions.

References

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys. K1 (1921) 966.
[2] O. Klein, Z. Phys. 37 (1926) 895.
[3] J. Scherk and J.H. Schwarz, Nucl. Phys. B81 (1974) 118;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
J.F. Luciani, Nucl. Phys. B135 (1978) 111;
E. Wissen, Nucl. Phys. B186 (1981) 412.
[4] F. Englert, Phys. Lett. 119B (1982) 339;
M.J. Duff and C.N. Pope, in: Supersymmetry and supergravity 82, eds. S. Ferrara, J.G. Taylor and P. van Nieuwenhuizen (World Scientific, Singapore, 1983).
[5] P. van Nieuwenhuizen, Phys. Rep. 68 (1981) 189.
[6] J.H. Schwarz, Phys. Rep. 89 (1982) 223.
[7] G. Chapline and R. Slansky, Nucl. Phys. B209 (1982) 461.
[8] R. Cowsik and J. McClelland, Phys. Rev. Lett. 29 (1972) 669.
[9] A. Salam and J. Strathdee, Ann. Phys. (NY) 141 (1982) 316.
[10] G. Steigman, Ann. Rev. Nucl. Part. Sci. 29 (1979) 313.
[11] J. Preskill, Phys. Rev. Lett. 43 (1979) 1365.
[12] S. Wolfram, Phys. Lett. 82B (1979) 65.
[13] Ya.B. Zeldovich and M. Khlopov, Phys. Lett. 79B (1978) 239.
[14] D.A. Dicus, D. Page and V.L. Teplitz, Phys. Rev. D26 (1982) 1306.
[15] See, e.g., S. Weinberg, Gravitation and cosmology (Wiley, New York, 1972).
[16] A. Chodos and S. Detweiler, Phys. Rev. D21 (1980) 2167.
[17] C. Wetterich, Phys. Lett. 113B (1982) 377.
[18] V.A. Rubakov and M.E. Shaposhnikov, Phys. Lett. 125B (1983) 139.
[19] R. Slansky, Phys. Rep. 79 (1981) 1.
[20] R. D'Auria and P. Fre, Phys. Lett. 121B (1983) 141.
[21] M.J. Duff, B.E.W. Nilsson and C.N. Pope, Phys. Rev. Lett. 50 (1983) 2043.

IS COSMOLOGICAL DIMENSIONAL REDUCTION POSSIBLE?

Kei-ichi MAEDA
Research Institute for Fundamental Physics, Kyoto University, Kyoto 606, Japan

Received 7 October 1983

Chodos and Detweiler have shown that "extra" dimensions in Kaluza–Klein theories may contract to an unobservable scale, by adopting an anisotropic Kasner-type vacuum solution of higher dimensions. However, if we consider a quantized matter field, the above anisotropic space–time may be isotropized by particle creation in the same manner with the process proposed by Zel'dovich in a conventional four-dimensional space–time. In this paper, the isotropization process in higher dimensions is computed by generalizing the method given by Hu and Parker for a conventional universe. It would be concluded that cosmological dimensional reduction does not occur.

The higher-dimensional theory of Kaluza and Klein is one of the most interesting ways of unifying gauge theories and gravitation. An important question in this approach is how the scale of the "extra" dimensions shrinks to be so small that one cannot observe it. Such a process must exist if we take the point of view that the "extra" dimension is not simply a mathematical stratagem but rather a physical reality. From the point of view of the solutions of the higher-dimensional Einstein equations, the methods of this dimensional reduction may be classified into two types; one is the static solution and the other is the dynamical one. In the latter case, we will call it cosmological dimensional reduction, which has been proposed first by Chodos and Detweiler [1], and in a more realistic model, i.e. eleven-dimensional supergravity theory, by Freund and Rubin [2,3]. These cosmological dimensional reductions are very attractive because one can explain the smallness of the "extra" dimensions by the dynamical evolution of the universe.

The vacuum Einstein equations have the Kasner-type solution, which describes an anisotropic and homogeneous space. Chodos and Detweiler have shown that the space of the "extra" dimension contracts to a very small scale by adopting the Kasner-type solution. For example, in the case of five dimensions, the scale factor of a conventional three-space expands as $a(t) \propto t^{1/2}$ and that of the fifth dimensional space contracts as $a_5(t) \propto t^{-1/2}$.

In the conventional four-dimensional Einstein theory, it is known that if we consider a quantum effect of matter field the particle creation mechanism isotropizes the anisotropic expansion. This isotropization process was first proposed by Zel'dovich [4] and the quantitative analysis has been given by Zel'dovich and Starobinsky [5] and Hu and Parker [6].

The model given by Chodos and Detweiler is a Kasner-type anisotropic universe in higher dimensions. Then, if we take into consideration the matter lagrangian, which is requested in more realistic unified theories, the above isotropization mechanism due to particle creation may destroy the cosmological dimensional reduction process. In the model by Freund and Rubin, the eleven-dimensional universe is also anisotropic and the matter lagrangian of an antisymmetric tensor field $A_{\mu\nu\rho}$ has been included. Then, isotropization may occur also in this case.

In this paper, we consider a quantum effect of a scalar field in five-dimensional space–time as a concrete model and investigate how effectively particle creation works for the isotropization of the five-dimensional universe which begins with an anisotropic Kasner-type solution. In the case of much higher dimensions, we can easily give a similar discussion and probably get a qualitatively same result.

We write the five-dimensional Einstein equations as [‡1]

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 8\pi G \langle T_{\mu\nu} \rangle_{\text{reg}}, \tag{1}$$

where $\langle T_{\mu\nu} \rangle_{\text{reg}}$ is a vacuum expectation value of the regularized energy–momentum tensor of the matter field. We assume a generalized Bianchi type I universe in the five dimensions, which has the following form:

$$ds^2 = -dt^2 + \sum_{A=1}^{5} a_A^2(t) (dx^A)^2 . \tag{2}$$

The cosmological scale factors $a_A(t)$ are determined by eq. (1). It follows from eqs. (1) and (2) that the nonvanishing components of $\langle T_\mu{}^\nu \rangle$ are

$$\langle T_0{}^0 \rangle \equiv -\rho, \quad \langle T_A{}^B \rangle \equiv P_A \delta_A{}^B, \tag{3}$$

where ρ is an energy density and the P_A are principal pressures. The Einstein equations (1) take the form:

$$G_0{}^0 = \tfrac{1}{2} V^{-1/2} \left[\sum_A (\alpha'_A)^2 - \left(\sum_A \alpha'_A \right) \left(\sum_B \alpha'_B \right) \right]$$

$$= -8\pi G_5 \rho, \tag{4}$$

and

$$R_A{}^A = V^{-1/2} \left[\alpha''_A + \tfrac{3}{4} \alpha'_A \left(\sum_B \alpha'_B \right) \right]$$

$$= 8\pi G_5 (P_A - \tfrac{1}{3} T), \tag{5}$$

since

$$R_{\mu\nu} = 8\pi G_5 (T_{\mu\nu} - \tfrac{1}{3} g_{\mu\nu} T), \tag{1'}$$

with $T = T_\mu{}^\mu$. A prime denotes a derivative with respect to conformal time, which is defined by

$$\eta = \int^t V^{-1/4} dt', \tag{6}$$

and $\alpha_A = \ln a_A$ and $V = a_1 a_2 a_3 a_5$. From the Bianchi identity we get the conservation equation of the energy–momentum tensor; $\langle T^{\mu\nu} \rangle_{;\nu} = 0$, that is,

$$V^{-1} (\rho V)' + \sum_A \alpha'_A P_A = 0. \tag{7}$$

The conservation equation (7) guarantees the validity of the constraint equation (4), if the constraint equation is satisfied at initial time. Therefore, we use eq. (7) instead of eq. (4).

We assume the ordinary three-dimensional space to be isotropic as in the case of Chodos and Detweiler. i.e. $a = a_1 = a_2 = a_3$. Then, the trace of the energy–momentum tensor is

$$\langle T_\mu{}^\mu \rangle = -\rho + (3P + P_5) \equiv T(\eta), \tag{8}$$

where $P = P_1 = P_2 = P_3$. From eqs. (7) and (8), we can express the pressure P and P_5 as

$$P = \tfrac{1}{3} (\beta' - \alpha')^{-1} \{ \rho' + (\ln V)' \rho + \beta' [T(\eta) + \rho] \}, \tag{9}$$

$$P_5 = (\alpha' - \beta')^{-1} \{ \rho' + (\ln V)' \rho + \alpha' [T(\eta) + \rho] \}, \tag{10}$$

where $\alpha = \ln a$ and $\beta = \alpha_5 = \ln a_5$.

As the matter field we consider a quantized conformal massless scalar field. The lagrangian of the scalar field is

$$\mathcal{L}_{\text{M}} = -\tfrac{1}{2} \sqrt{-g} \, (g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + \xi R \phi^2), \tag{11}$$

where $\xi = \frac{3}{16}$ in the five-dimensional space–time. The energy–momentum tensor is given by

$$T_{\mu\nu} = (\nabla_\mu \phi)(\nabla_\nu \phi) - \tfrac{1}{2} g_{\mu\nu} (\nabla_\rho \phi)(\nabla^\rho \phi) + \tfrac{3}{16} \phi^2 G_{\mu\nu}$$

$$- \tfrac{3}{16} [\nabla_\mu \nabla_\nu (\phi^2) - g_{\mu\nu} \nabla^\rho \nabla_\rho (\phi^2)] . \tag{12}$$

$T_\mu{}^\mu$ vanishes on the classical level, but, in a quantum theory, a non-vanishing trace anomaly $T(\eta)$ may arise as a result of the regularization. We write the initial epoch when the particle creation mechanism switches on as t_0. Since in our calculations t_0 is greater than the Planck time t_{P}, we have ignored the trace anomaly and put $T(\eta) = 0$. So, if we can calculate the density $\rho = -\langle T_0{}^0 \rangle_{\text{reg}}$ and ρ', we get from eqs. (9) and (10) the source term which arises as a result of particle creation and we find the behavior of a and a_5 by solving (5).

Here, we shall estimate ρ and ρ' by the five-dimensional version of the method of Hu and Parker [6]. Let us put

$$\phi = V^{-3/8} \chi, \tag{13}$$

with

[‡1] Greek letters take values from 0 to 3 and 5, capital Latins from 1 to 3 and 5, and small Latins from 1 to 3. Our signature is the same as that of Misner et al. [7].

$$\chi = \frac{1}{(2\pi)^2} \int \mathrm{d}^4 k \, [A_{\boldsymbol{k}} \chi_{\boldsymbol{k}}(\eta) \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}) + A_{\boldsymbol{k}}^{\dagger} \chi_{\boldsymbol{k}}^{*}(\eta) \exp(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})] . \tag{14}$$

The function $\chi_{\boldsymbol{k}}(\eta)$ satisfies the equation

$$\chi_{\boldsymbol{k}}'' + (\Omega_{\boldsymbol{k}}^2 + Q)\chi_{\boldsymbol{k}} = 0, \tag{15}$$

where

$$\Omega_{\boldsymbol{k}} = V^{1/4} \left(\sum_A \frac{k_A^2}{a_A^2} \right)^{1/2} \tag{16}$$

and

$$Q = \frac{3}{64} \sum_{A>B} (\alpha_A' - \alpha_B')^2 . \tag{17}$$

We can show that the operators $A_{\boldsymbol{k}}$ and $A_{\boldsymbol{k}}^{\dagger}$ obey the usual commutation relations of annihilation and creation operators, respectively, using the relations of canonical quantization provided that

$$\chi_{\boldsymbol{k}}^{*\prime} \chi_{\boldsymbol{k}} - \chi_{\boldsymbol{k}}^{*} \chi_{\boldsymbol{k}}' = \mathrm{i}. \tag{18}$$

Let $|0_A\rangle$ be a vacuum state defined by the annihilation operators $A_{\boldsymbol{k}}$. Adopting the adiabatic regularization proposed by Fulling et al. [8], we can write down the regularized energy density as follows:

$$\begin{aligned} \rho &= -\langle 0_A | T_0{}^0 | 0_A \rangle_{\mathrm{reg}} \\ &= \frac{1}{32\pi^4} V^{-5/4} \int \mathrm{d}^4 k \{ |\chi_{\boldsymbol{k}}'|^2 + (\Omega^2 - Q)|\chi_{\boldsymbol{k}}|^2 \\ &\quad - \Omega - (1/2\Omega)[\tfrac{1}{4}(\Omega'/\Omega)^2 - Q] \\ &\quad + (1/2\Omega)[\tfrac{1}{8}(\Omega'/\Omega)^2 \epsilon_{2(2)} - \tfrac{1}{4}(\Omega'/\Omega)\epsilon_{2(3)}' \\ &\quad - \tfrac{1}{4}\Omega^2(\epsilon_{2(2)})^2 - \tfrac{1}{2} Q \epsilon_{2(2)}]\}, \end{aligned} \tag{19}$$

where the definitions of $\epsilon_{2(2)}$ and $\epsilon_{2(3)}$ are the same as in eqs. (2.40) and (2.41) in ref. [8] and $\Omega = \Omega_{\boldsymbol{k}}$.

Following Hu and Parker, we decompose the energy density ρ into a quantum part $\rho_{q(t)}$ and a classical part $\rho_{c(t)}$, where $q(t)$ is the quantum domain in $\boldsymbol{k}$-space over which $\omega_{\boldsymbol{k}} (\equiv V^{-1/4}\Omega_{\boldsymbol{k}}) \lesssim t^{-1}$ and $c(t)$ is the classical one with $\omega_{\boldsymbol{k}} > t^{-1}$. In the low frequency domain $q(t)$ of the $\boldsymbol{k}$-space, quantum effects such as particle creation are dominant and for the higher frequency region $c(t)$ where the WKB approximation is valid, matter can be treated as classical incoherent matter. For low frequencies, i.e. in the domain $q(t)$, we get the solution of eq. (16) using the approximation $\exp(2\mathrm{i} \int_{\eta_0}^{\eta} \Omega_{\boldsymbol{k}} \, \mathrm{d}\eta') \approx 1$ as follows:

$$\chi_{\boldsymbol{k}} = \frac{1}{\sqrt{2\Omega}} \left[\alpha_{\boldsymbol{k}} \exp\left(-\mathrm{i} \int_{\eta_0}^{\eta} \Omega \, \mathrm{d}\eta' \right) + \beta_{\boldsymbol{k}} \exp\left(\mathrm{i} \int_{\eta_0}^{\eta} \Omega \, \mathrm{d}\eta' \right) \right], \tag{20}$$

$$\begin{aligned} \alpha_{\boldsymbol{k}} &= c_1 \left(\Omega^{1/2} - \mathrm{i}\Omega^{-1/2} \int_{\eta_0}^{\eta} Q \, \mathrm{d}\eta' \right) + c_2 \Omega^{-1/2}, \\ \beta_{\boldsymbol{k}} &= c_1 \left(\Omega^{1/2} + \mathrm{i}\Omega^{-1/2} \int_{\eta_0}^{\eta} Q \, \mathrm{d}\eta' \right) - c_2 \Omega^{-1/2}, \end{aligned} \tag{21}$$

where c_1 and c_2 are integral constants [6].

We take the initial condition such that $\alpha_{\boldsymbol{k}}(\eta_0) = 1$ and $\beta_{\boldsymbol{k}}(\eta_0) = 0$, when the initial vacuum state $|0_A\rangle$ with respect to $A_{\boldsymbol{k}}$ is regarded as a real vacuum because $\chi_{\boldsymbol{k}}$ becomes a positive-frequency solution. From this condition we find,

$$c_1 = \tfrac{1}{2}\Omega_{\boldsymbol{k}}(\eta_0)^{-1/2}, \quad c_2 = \tfrac{1}{2}\Omega_{\boldsymbol{k}}(\eta_0)^{1/2}. \tag{22}$$

Then we can compute the quantum part of the energy density $\rho_{q(t)}$ from eqs. (19)–(22) as

$$\begin{aligned} \rho_{q(t)} &= \frac{1}{32\pi^4} V^{-5/4} \int_{q(t)} \mathrm{d}^4 k \left\{ \frac{1}{2\Omega_0} \left[(\Omega^2 - Q) + \left(\int_{\eta_0}^{\eta} Q \, \mathrm{d}\eta' \right)^2 \right] + \tfrac{1}{2}\Omega_0 \right. \\ &\quad \left. - \Omega - (1/2\Omega)[\tfrac{1}{4}(\Omega'/\Omega)^2 - Q] + (1/2\Omega)[\ldots] \right\}, \end{aligned} \tag{23}$$

where

$$\Omega_0 = \Omega(\eta_0).$$

Since the quantum domain $q(t)$ shrinks with time, some part of $\rho_{q(t)}$ changes into $\rho_{c(t)}$. Let $\boldsymbol{k}_m = (\boldsymbol{k}_m(t), k_{5m}(t))$ be the maximum wave number defined by $\omega_{\boldsymbol{k}}(t) = t^{-1}$ with $k \equiv (\Sigma_i k_i^2)^{1/2}$. The part of the energy density which becomes classical in the time interval from η to $\eta + \Delta\eta$ is estimated as

$$\delta\rho_q(\eta) = \rho_q(\boldsymbol{k}_m(\eta), \eta) - \rho_q(\boldsymbol{k}_m(\eta + \Delta\eta), \eta). \tag{24}$$

We assume that the classical particles in the domain

$c(t)$ behave as a collision-dominated relativistic fluid with "isotropic" pressure $P_c(=\frac{1}{4}\rho_c)$, and then $\rho_c(t)$ is proportional to $V^{-5/4}$. This assumption may be regarded as a phenomenological way of taking into account the effect of strong interactions occurring in the high-density fluid of newly created particles [6]. Then, the density at $\eta + \Delta\eta$ is given by

$$\rho(\eta + \Delta\eta) = [V(\eta)/V(\eta + \Delta\eta)]^{5/4} [\rho_c(\eta) + \delta\rho_q(\eta)]$$

$$+ \rho_q(\boldsymbol{k}_m(\eta + \Delta\eta), \eta + \Delta\eta)$$

$$= [V(\eta)/V(\eta + \Delta\eta)]^{5/4} \rho(\eta)$$

$$+ \Delta\eta [(\partial/\partial\eta)\rho_q(\boldsymbol{k}_m, \eta) + \tfrac{5}{4}(V'/V)\rho_q(\boldsymbol{k}_m, \eta)] . \quad (25)$$

Giving the initial energy density, we can compute $\rho(\eta)$ and $\rho'(\eta)$ from eqs. (23) and (25), and then $P(\eta)$ and $P_5(\eta)$ follow from eqs. (9) and (10). We have taken our initial state to be vacuum. Then, through the particle creation process, the quantum energy density is created at first, and the particle whose frequency is larger than t^{-1} begins to behave like a classical particle. By the vacuum polarization effect, $\rho_q(\eta)$ in eq. (23) does not vanish initially. The regularization terms in $\rho_q(\eta)$, i.e. the last three terms of eq. (23) have been introduced in order to subtract the UV divergence by the adiabatic regularization. However, they will not play an important role for low frequency modes in the domain $q(t)$. Therefore, we will neglect all these terms in this paper, and we take only first two terms in eq. (23) ‡2.

We choose the initial condition as follows:

(i) $\rho(t_0) = \rho_{q(t_0)}$ and $\rho'(t_0)$ are given by eqs. (23) and (25). The initial energy density $\rho_{q(t_0)}$ does not vanish because of the vacuum polarization.

(ii) We cannot put $\alpha'(\eta_0) = 2/3\eta_0$ and $\beta'(\eta_0) = -2/3\eta_0$ as in the Kasner vacuum because this condition is inconsistent with the constraint equation (4) in the case of non-vanishing energy density. We choose $\alpha'(\eta_0) = 2/3\eta_0$ and $\beta'(\eta_0)$ is determined from eq. (4). $\beta'(\eta_0)$ is nearly equal to $-2/3\eta_0$, since $\rho_{q(t_0)}$ is not so large.

We show the result by the numerical integration

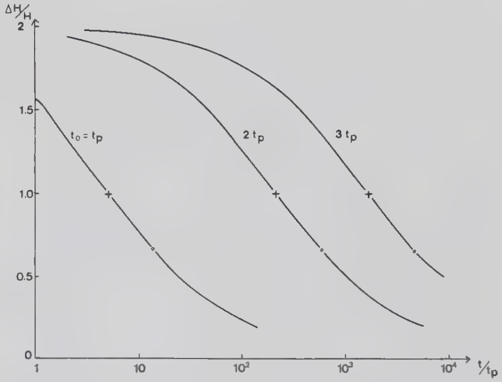


Fig. 1. A measure of anisotropy, $\Delta H/H$, is plotted with respect to t/t_P for the initial time $t_0 = t_P$, $2t_P$ and $3t_P$. + and ○ denote the characteristic isotropization time t_F and the time when the fifth dimension turns from contraction to expansion, respectively.

of eqs. (5) in fig. 1. The five-dimensional gravitational constant is chosen as $G_5 = 2\pi G^{3/2}$ because $G_5 \sim 2\pi G R_5$ and $R_5 \gtrsim l_P$, where R_5 is the present size of the fifth dimension and l_P is the Planck length. We define a measure of anisotropy by

$$\Delta H/H \equiv [(\alpha' - \beta')/2] \Big/ \left[\left(\sum_A \alpha'_A \right) \Big/ 4 \right] .$$

In a spatially flat universe, we can freely scale the axes a_A. For convenience, we set $a = a_5 = 1$ at t_0. From fig. 1, we can see that the isotropization mechanism by the particle creation is appreciably effective. We define the characteristic isotropization time t_F as the time when $\Delta H/H$ becomes unity. If we take $t_0 = t_P$, t_F is about $5t_P$ and the epoch when the fifth dimension turns from contraction to expansion is about $14t_P$. When $t_0 = 2t_P$ and $t_0 = 3t_P$, $t_F \sim 200t_P$ and $t_F \sim 1700t_P$ respectively. If we take the value of R_5 larger than l_P, e.g. $R_5 \sim 10l_P$, G_5 becomes greater than $2\pi G^{3/2}$ and then isotropization occurs more rapidly because the effect by the matter is enhanced as can be seen from eq. (5). Therefore, if we consider the quantum effect of a matter field in an anisotropic space–time like a Kasner-type solution, space–time is isotropized in five dimensions by the particle creation mechanism, and the cosmological dimensional reduction by Chodos and Detweiler may break down. If the initial time t_0 is before the Planck time, the above isotropization mechanism becomes more effective. In this case, however, since the trace anomaly term, $T(\eta)$, may become important, we cannot say anything within the framework of this paper.

---

‡2 Even if we take into account the regularization terms adopting the other regularization which is valid in the low frequency region, our result will not change qualitatively as far as the created energy density is positive.

Chodos and Detweiler consider that the five-dimensional space–time is closed in order to compactify the fifth dimension, i.e., $0 \leqslant x^A \leqslant L$. In the above calculation, we assume the space–time to be not closed. If $L \gg l_P$, the low frequency region $q(t)$ of $\boldsymbol{k}$-space contains many quantized waves because the critical wave number is about $t_0^{-1} \sim l_P^{-1}$ and the minimum wave number is about L^{-1}, and therefore we can turn Σ_{k_n} into $\int \mathrm{d}^4 k$ approximately. Therefore, even when space–time is closed, if $L \gg l_P$, the above result is valid. When $L \sim l_P$, the result may change quantitatively, but not qualitatively.

As an other effect of the closed space, we can consider the Casimir effect. Recently, Appelquist and Chodos [9] show that the five-dimensional static Kaluza–Klein vacuum ($M^4 \times S^1$) is unstable for a contraction of the fifth dimension when we consider a quantum effect of the gravitational field, calculating the effective potential. Therefore, we may be able to explain the smallness of the "extra" dimensions by this mechanism. We can consider that this instability is due to the Casimir effect because the fifth dimension is closed. Does the Casimir effect work also sufficiently in time-dependent background space–time like a universe? This effect may compete against the isotropization of space–time by particle creation.

Which mechanism is more important and effective, particle creation of matter fields or the Casimir effect of matter fields and/or gravitons? It is also not known whether the creation of gravitons isotropizes space–time, because gravity is not a conformally invariant theory. We must investigate the quantum effect and/or the topological effect of the gravitational field and the matter field in anisotropic time-dependent space–time.

Freund and Rubin also give a model of cosmological dimensional reduction in a more realistic theory. This higher dimensional universe is also anisotropic. In this model, the matter field, $A_{\mu\nu\rho}$, is included in the lagrangian. Is the anisotropic space–time also isotropized by the creation of this particle? The difference from the model by Chodos and Detweiler is that the existence of the anisotropic classical matter field is assumed to achieve compactification. Then, it is not so clear whether the mechanism of isotropization by quantum particle creation is effective, because the anisotropic matter may prevent the isotropization of space–time. However, this anisotropic classical matter give a non-zero cosmological constant and must vanish sooner or later. If the anisotropic matter vanishes near the Planck time and the universe is time-dependent, the particle creation mechanism may also isotropize the universe in eleven-dimensional space–time.

The author would like to thank Professor H. Sato for valuable discussions and continuous encouragement. He also thanks S. Midorikawa and U. Carow for useful discussions.

References

[1] A. Chodos and S. Detweiler, Phys. Rev. D21 (1980) 2167.
[2] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233.
[3] P.G.O. Freund, Nucl. Phys. B209 (1982) 146.
[4] Ya.B. Zel'dovich, Sov. Phys. JETP Lett. 12 (1970) 307.
[5] Ya.B. Zel'dovich and A.A. Starobinsky, Sov. Phys. JETP 34 (1971) 1159.
[6] B.L. Hu and L. Parker, Phys. Rev. D17 (1978) 933.
[7] C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (Freeman, San Francisco, 1973).
[8] S.A. Fulling, L. Parker and B.L. Hu, Phys. Rev. D10 (1974) 3905.
[9] T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141; Phys. Rev. D28 (1983) 772.

ON KALUZA–KLEIN COSMOLOGY

S. RANDJBAR-DAEMI [1]
International Centre for Theoretical Physics, Trieste, Italy
and Institute for Theoretical Physics, University of Bern, Bern, Switzerland

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England

and

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 16 November 1983

A generalization of Friedmann–Robertson -Walker cosmology to $4 + K$ dimensions is considered. The space–time manifold $R^1 \times S^3 \times S^K$ is characterized by two time-dependent scales, $R(t)$ and $a(t)$. The equations of motion for R and a are derived from the $4 + K$-dimensional Einstein action supplemented by a one-loop thermal term, corresponding to a gas of non-interacting scalar particles. It is shown that in the approximation when $T < 1/a$ the equations of motion admit a solution in which the internal space has a constant radius a while the external $R(t)$ evolves in the usual manner.

1. Einstein equations. The idea of an internal space whose symmetries correspond to the observed internal symmetries of low energy physics has received much attention, recently. It is assumed in this picture that space–time has more than four dimensions and that all but four of them are highly curved and compactified so as to be unresolvable with the energies available at present. The relevant distance (measured in units of Planck length) would be inversely correlated with the gauge coupling constants in the theory.

As with other questions involving inaccessibly small distance scales it seems appropriate to look for cosmological implications of the Kaluza–Klein hypothesis. Were the compact dimensions always compact and always so much more curved than the others? If there is a temporal evolution of the distance scales, is it compatible with the observed constancy of gauge coupling parameters like the fine structure constant? We shall demonstrate that it is indeed possible to maintain a mixed internal scale while the external radius evolves in the conventional manner.

To study the cosmological aspects [1] of the Kaluza–Klein hypothesis we shall begin by setting up a straightforward generalization of the homogeneous Friedmann, Roberston, Walker cosmologies for space–times with extra dimensions. To be specific we shall assume the geometry $R^1 \times S^3 \times S^K$ where R^1 corresponds to the time dimension, S^3 to the "large" space dimensions and S^K to the "small" or internal space dimensions. Instead of the sphere S^3 one could choose an open space of constant curvature and, instead of the sphere S^K one could choose a compact quotient space G/H of K dimensions. In any case the metric tensor could be presented in the form

$$g_{MN} = \begin{pmatrix} -1 & & \\ & R(t)^2 \tilde{g}_{mn}(x) & \\ & & a(t)^2 \tilde{g}_{\mu\nu}(y) \end{pmatrix} , \quad (1.1)$$

---

[1] Supported by Schweizerischer National Funds.

where $\tilde{g}_{mn}(x)$ and $g_{\mu\nu}(y)$ represent metrics on the three- and K-dimensional spaces, respectively. The sizes of these two spaces are characterized by time-dependent scale factors, $R(t)$ and $a(t)$.

The metric (1.1) admits a number of isometries. For the case $\mathrm{R}^1 \times \mathrm{S}^3 \times \mathrm{S}^K$, the symmetry group of the group of the spatial section is $\mathrm{O}(4) \times \mathrm{O}(K+1)$. The stress tensor, whose form would be dictated by Einstein's equations, must have the same invariances. In the coordinate system where the metric is given by (1.1), the stress tensor is

$$\mathsf{T}_{MN} = \begin{bmatrix} \rho & & \\ & p g_{mn} & \\ & & p' g_{\mu\nu} \end{bmatrix} , \qquad (1.2)$$

where the energy density ρ and pressures p, p' may depend on time, but not on the space coordinates x^m, y^μ.

The Einstein equations for the scale factors R and a in terms of ρ, p and p' can be derived from the action functional,

$$\Gamma = \int \mathrm{d}^{4+K}x(-g)^{1/2}[-G^{-1}(\mathsf{R} + \bar{\Lambda})] + \Gamma_1(g) , \quad (1.3)$$

where R denotes the $4+K$-dimensional Ricci scalar, $\bar{G}$ is the coupling parameter and $\bar{\Lambda}$ is a cosmological constant. The source term, Γ_1, represents the contributions of matter fields, vacuum fluctuations, etc. Its variation defines the stress tensor,

$$\delta\Gamma_1 = \int \mathrm{d}^{4+K}x(-g)^{1/2}\,\tfrac{1}{2}\delta g^{MN}\mathsf{T}_{MN} . \qquad (1.4)$$

By substituting the $\mathrm{O}(4) \times \mathrm{O}(K+1)$ invariant ansatz (1.1) into (1.3), one can reduce this action to the form

$$\Gamma = \int \mathrm{d}t\, \mathcal{L}(R, a) ,$$

where the lagrangian, $\mathcal{L}$, is given by

$$\mathcal{L} = -(\Omega_3\Omega_K/\bar{G})$$
$$\times\ [6\dot{R}^2/R^2 + 6K(\dot{R}/R)\,\dot{a}/a + K(K-1)\dot{a}^2/a^2] - U. \qquad (1.5)$$

Here Ω_3 and Ω_K denote the spatial volumes,

$$\Omega_3 = 2\pi^2 R^3, \quad \Omega_K = (2\pi)^{(K+1)/2} a^K / \Gamma((K+1)/2) \qquad (1.6)$$

and U is an effective "potential". The construction of this potential will be discussed more fully in the following section. It incorporates a classical gravity term arising from $\mathsf{R} + \Lambda$ [and displayed in the first bracket of (1.7)] plus terms in the second bracket which arise from Γ_1. The latter has a (one-loop) quantum part plus a thermal part depending on entropy, S

$$U = (\Omega_3\Omega_K/\bar{G})[-6/R^2 - K(K-1)/a^2 + \bar{\Lambda}]$$
$$+ (\sigma_K\Omega_3/a^4 + \tau S^{4/3}R + \zeta m) . \qquad (1.7)$$

Here σ_K, τ and ζ are numerical parameters which can be computed. The approximations which go into the form of the entropy (S)-dependent term in (1.7) will be discussed in the next section, but we believe that the given expression should be valid in the regime

$$R \gg 1/T \gg a , \qquad (1.8)$$

where the temperature T is given by

$$T = \partial U/\partial S \sim S^{1/3}/R . \qquad (1.9)$$

The time evolution of the Kaluza–Klein universe is governed by the Euler–Lagrange equations derived from (1.5). We now show that for an appropriate choice of the parameter $\bar{\Lambda}$ in U, it is possible to obtain a solution where the internal radius is constant ($\dot{a} = 0$) while the large radius evolves as in standard model cosmology.

The two equations of motion simplify, on setting $\dot{a} = 0$, to the following:

$$\partial_t(R\dot{R}) - \tfrac{1}{2}\dot{R}^2 = \tfrac{1}{12}\bar{G}(R^2/\Omega_3\Omega_K)R\,\partial U/\partial R , \quad (1.10a)$$

$$R^{-1}\partial_t(R^2\dot{R}) - \dot{R}^2 = (\bar{G}/6K)(R^2/\Omega_3\Omega_K)a\partial U/\partial a . \qquad (1.10b)$$

The "energy" integral corresponding to these equations reduces, when $\dot{a} = 0$, to

$$E = -(\Omega_3\Omega_K/\bar{G})(6\dot{R}^2/R^2) + U . \qquad (1.11)$$

The compatibility of (1.10a), (1.10b) and (1.11) requires the algebraic relation

$$R\,\partial U/\partial R - (2/K)a\partial U/\partial a = E - U . \qquad (1.12)$$

On taking $E = \zeta m$, this condition reduces to

$$f(a) = \mathrm{d}f/\mathrm{d}a = 0 , \qquad (1.13a)$$

where [‡1]

[‡1] Eqs. (13a) are the same as those of Candelas and Weinberg [2] for the static case considered by them ($R = \infty$).

$$f(a) = (\Omega_K/\bar{G})[-K(K-1)/a^2 + \bar{\Lambda}] + \sigma/a^4 \ . \qquad (1.13b)$$

Determining $\bar{\Lambda}$ and a in terms of $\bar{G}$ from (1.13a) and substituting the results into (1.10b), we obtain,

$$\partial_t(R\dot{R}) = -1 \ , \qquad (1.14)$$

which is solved by

$$R^2 = t(t_0 - t) \ , \qquad (1.15)$$

whre t_0, the time of collapse, is given by (1.11), i.e. $t_0^2 = (\tau/3\pi^2)(\bar{G}/\Omega_K)S^{4/3}$.

The expression (1.15) for $R^2(t)$ gives the conventional Friedmann–Robertson–Walker cosmology for the case when the standard three-curvature parameter k equals unity. In the general case, the $6/R^2$ term in the first bracket in (1.7) reads $6k/R^2$ and the right-hand side of (1.14) equals $(-k)$, with appropriate changes in (1.15).

To summarize, if $R \gg 1/T > a$, and if $\bar{\Lambda}$ and a are determined in terms of $\bar{G}$ from (1.13a), we find $\dot{a}(t) = 0$, while $R(t)$ evolves in the conventional manner.

2. The free energy. In order to have a meaningful expression for the components of the stress tensor, it is necessary to derive them from a free energy function, $F(R, a, T)$. From F one obtains the energy as a function of the entropy, $S = -\partial F/\partial T$, by the usual Legendre transform

$$U = F + TS = \Omega_3\Omega_K\rho(R, a, S) \ . \qquad (2.1)$$

If the pressures p and p' of eq. (1.2) are defined by the thermodynamic identity,

$$\mathrm{d}(\Omega_3\Omega_K\rho) = T\mathrm{d}S - \Omega_K p\, \mathrm{d}\Omega_3 - \Omega_3 p'\, \mathrm{d}\Omega_K \ , \qquad (2.2)$$

then as is well-known the conservation of energy, $\nabla_M T^M{}_N = 0$, implies conservation of entropy. It follows that the Einstein equations deriving from (1.5) require that S be independent of time. This was tacitly assumed in section 1.

Since we have no a priori idea of what equations of state would be appropriate for Kaluza–Klein cosmology, we shall proceed on the basis of a simple model. We shall suppose that the density and pressures are due to a gas of non-interacting, spinless particles in thermal equilibrium. The free energy of such a system can be expressed by a one-loop integral,

$$\beta F = \tfrac{1}{2} \ln \mathrm{Det}(-\Box + \mu^2) \ , \qquad (2.3)$$

where μ is a mass parameter and $\Box$ denotes the laplacian on the compact manifold $S^1 \times S^3 \times S^K$. This manifold is characterized by three radii, $\beta/2\pi = 1/2\pi T$, R and a. The expression (2.3) is given formally by a triple sum over the eigenvalues of $-\Box + \mu^2$,

$$\beta F = \frac{1}{2} \sum_{r=-\infty}^{\infty} \sum_{m,n=0}^{\infty} D_{mn} \times \ln[(2\pi r/\beta)^2 + m(m+2)/R^2 + n(n+K-1)/a^2 + \mu^2] \ , \qquad (2.4)$$

where D_{mn} is a multiplicity factor. It is equal to the dimension of the $O(4) \times O(K+1)$ representation $(m, 0) \times (n, 0, ..., 0)$ in Gel'fand Zetlin notation,

$$D_{mn} = (m+1)^2 \times (2n + K - 1)(n + K - 2)!/(K-1)!n! \ . \qquad (2.5)$$

The sum (2.4) is of course divergent. But it is easy to regularize [3]. For this purpose write

$$\ln X = \frac{\mathrm{d}}{\mathrm{d}s} X^s |_{s=0} = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{1}{\Gamma(-s)} \int_0^\infty \mathrm{d}t\, t^{s-1} \exp(-tX)\right)_{s=0}$$

for each term in the sum. After some rearrangements one obtains an expression for the finite part

$$\beta F = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{1}{2\Gamma(-s)} \int_0^\infty \mathrm{d}t\, t^{-s-1} \exp(-t\mu^2) \times \sigma_1(4\pi^2 t/\beta^2)\sigma_3(t/R^2)\sigma_K(t/a^2)\right)_{s=0} \ , \qquad (2.6)$$

where the functions σ_1, σ_3 and σ_K are defined, respectively, by

$$\sigma_1(4\pi^2 t/\beta^2) = \sum_{-\infty}^{\infty} \exp[-t(2\pi r/\beta)^2] \ ,$$

$$\sigma_3(t/R^2) = \sum_0^{\infty} (m+1)^2 \exp[-m(m+2)t/R^2] \ ,$$

$$\sigma_K(t/a^2) = \sum_0^{\infty} [(2n + K - 1)(n + K - 2)!/(K-1)!n!] \times \exp[-n(n+K-1)t/a^2] \ . \qquad (2.7)$$

These sums converge and define analytic functions on the half-plane Re $t > 0$. At $t = 0$ they are singular,

$$\sigma_1 \sim t^{-1/2}, \quad \sigma_3 \sim t^{-3/2}, \quad \sigma_K \sim t^{-K/2} .$$

To evaluate the integral (2.6) it is necessary to make approximations to the σ_i. The most important observation in this connection [4] is that these functions are closely related to the Jacobi Theta functions [5]. In particular,

$$\theta_3(0|\mathrm{i}u/\pi) = \sum_{-\infty}^{\infty} \exp(-r^2u)$$

$$= \sqrt{\pi/u} \sum_{-\infty}^{\infty} \exp[-(\pi r)^2 u^{-1}] . \tag{2.8}$$

The two series on the right-hand side provide highly convergent expansions for the σ_i around $u = \infty$ and around $u = 0$ [‡2], respectively. In particular, the zero-temperature limit corresponds to the term

$$\sigma_1 \simeq (4\pi t/\beta^2)^{-1/2} , \tag{2.9}$$

and gives

$$F(R, a, T = 0) = \frac{\mathrm{d}}{\mathrm{d}s}\Bigg(\frac{1}{4\sqrt{\pi}\,\Gamma(-s)}$$

$$\times \int_0^\infty \mathrm{d}t\, t^{-s-3/2} \exp(-t\mu^2)\sigma_3(t/R^2)\sigma_K(t/a^2)\Bigg).$$

In a similar fashion, the flat space limit, $R \to \infty$, of the energy density, $F/2\pi^2 R$, corresponds to

$$\sigma_3 \simeq \tfrac{1}{4}\sqrt{\pi} R^3 t^{-3/2} . \tag{2.10}$$

The effective potential of section 1 was obtained by assuming $R \gg \beta > a$, while the mass parameter μ for the scalar particles is small (more precisely $\mu < T$). To derive it, use (2.9) and (2.11) to obtain,

$$F \simeq \frac{R^3}{16} \frac{\mathrm{d}}{\mathrm{d}s}\Bigg[\frac{1}{\Gamma(-s)} \int_0^\infty \mathrm{d}t \;\; t^{-s/3}$$

$$\times \left(1 + 2\sum_1^\infty \exp(-r^2\beta^2/4t)\right) \sigma_K(t/a^2)\Bigg]_{s=0}$$

$$= \frac{R^3}{16}\Bigg[\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{a^{-4-2s}}{\Gamma(-s)} \int_0^\infty \mathrm{d}u\, u^{-s-3} \sigma_K(u)\right)_{s=0}$$

$$- 2\int_0^\infty \mathrm{d}t\, t^{-3} \sum_1^\infty \exp(-r^2\beta^2/4t)\sigma_K(t/a^2)\Bigg] . \tag{2.11}$$

The "s-regularization" can be safely removed in the temperature dependent terms, which converge at $t = 0$. It can also be removed in the zero-temperature term if K is odd, since the integral involving $\sigma_K(u)$ is analytic at $s = 0$ in that case. (For even K one obtains $\sim a^{-4} \ln a$ in the limit $s = 0$.) Thus

$$F \simeq \Omega_3(\sigma_K a^{-4} - \tfrac{1}{90}\pi^2\beta^{-4} + ...) , \tag{2.12}$$

where

$$\sigma_K = -\frac{1}{32\pi^2} \int_0^\infty \mathrm{d}u\, u^{-3} \sigma_K(u) . \tag{2.13}$$

In obtaining the second term of (2.12) we have taken $\sigma_K = 1$, in effect discarding the exponentially small terms $\sim\exp(-\beta/a)$. (Otherwise stated, only the internal space zero modes are retained in the sum over states.) Details of these various approximations will be discussed elsewhere.

The one-loop contribution to the energy, used in section 1, is then obtained by taking the Legendre transform of (2.12)

$$U \simeq \sigma_K \Omega_3/a^4 + \tau S^{4/3}/R , \tag{2.14}$$

where

$$\tau = \tfrac{3}{4}(45/4\pi^4)^{1/3} . \tag{2.15}$$

[If there are other species of particles, in addition to scalars which circulate around the loop (for example gravitons, gauge particles, or spinors) the numerical values of the parameters σ_K and τ will change.]

3. Conclusions. It is interesting to confirm that the

[‡2] This seems to be true for the odd-dimensional spheres. For the even-dimensional spheres one probably has to employ asymptotic series $\sigma_K(u^{-K/2}(\alpha + \beta u + \gamma u^2 + ...)$, near $u = 0$.

Kaluza–Klein cosmology does admit of a time-independent internal radius a (consistent with lack of variability of gauge couplings with time), in an approximation to the free energy where $R \gg 1/T > a$, while $\bar{\Lambda}$ and a are suitably correlated by (1.13a). For the regime of temperatures where our *precise* approximation for the thermal contribution to the free energy, $(S^{4/3}/R)$, may be expected to break down – for example for planckian temperatures, $T \gtrsim 1/a$ – phase transitions [2] in the viour of a (and R) may occur. These will considered elsewhere.

The authors would like to thank Professor L Van Hove for a discussion.

References

[1] P.G.O. Freund, Nucl. Phys. B209 (1982) 146;
T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141;
E. Alvarez and M. Belen Gavela, Phys. Rev. Lett. 51 (1983) 931;
M. Rubin and B.D. Roth, Phys. lett. 127B (1983) 55; Nucl. Phys. B226 (1983) 44;
T. Appelquist, A. Chodos and E. Meyers, Yale preprint YPT 83-04 (1983);
Q. Shafi and C. Wetterich, Phys. Lett. 129B (1983) 387.

[2] P. Candelas and S. Weinberg, Texas preprint UTTG-8-83 (1983).

[3] A. Salam and J.A. Strathdee, Nucl. Phys. B90 (1975) 203.

[4] C.M. Mukku, J.A. Magpantay and W.A. Sayed, Ann. Phys. (NY) 145 (1983) 27.

[5] A. Erdelyi et al., eds., Higher transcendental functions (McGraw-Hill, New York, 1953) Vol. II.

CHAPTER

7

Quantum Effects

On the attraction between two perfectly conducting plates

BY

H. B. G. CASIMIR

(Communicated at the meeting of May 29, 1948.)

In a recent paper by POLDER and CASIMIR [1]) it is shown that the interaction between a perfectly conducting plate and an atom or molecule with a static polarizibility α is in the limit of large distances R given by

$$\delta E = -\frac{3}{8\pi} \hbar c \frac{\alpha}{R^4}$$

and that the interaction between two particles with static polarizibilities α_1 and α_2 is given in that limit by

$$\delta E = -\frac{23}{4\pi} \hbar c \frac{\alpha_1 \alpha_2}{R^7}.$$

These formulae are obtained by taking the usual VAN DER WAALS–LONDON forces as a starting point and correcting for retardation effects.

In a communication to the "Colloque sur la théorie de la liaison chimique" (Paris, 12—17 April, 1948) the present author was able to show that these expressions may also be derived through studying by means of classical electrodynamics the change of electromagnetic zero point energy. In this note we shall apply the same method to the interaction between two perfectly conducting plates.

Let us consider a cubic cavity of volume L^3 bounded by perfectly conducting walls and let a perfectly conducting square plate with side L be placed in this cavity parallel to the xy face and let us compare the situation in which this plate is at a small distance a from the xy face and the situation in which it is at a very large distance, say $L/2$. In both cases the expressions $\frac{1}{2}\Sigma\hbar\omega$ where the summation extends over all possible resonance frequencies of the cavities are divergent and devoid of physical meaning but the *difference* between these sums in the two situations, $\frac{1}{2}(\Sigma\hbar\,\omega)_I - \frac{1}{2}(\Sigma\hbar\,\omega)_{II}$, will be shown to have a well defined value and this value will be interpreted as the interaction between the plate and the xy face.

The possible vibrations of a cavity defined by

$$0 \leqslant x \leqslant L, \quad 0 \leqslant y \leqslant L, \quad 0 \leqslant z \leqslant a$$

have wave numbers

$$k_x = \frac{\pi}{L} n_x, \quad k_y = \frac{\pi}{L} n_y, \quad k_z = \frac{\pi}{a} n_z,$$

where n_x, n_y, n_z are positive integers;

$$k=\sqrt{k_x^2+k_y^2+k_z^2}=\sqrt{\varkappa^2+k_z^2}.$$

To every k_x, k_y, k_z correspond *two* standing waves unless one of the n_i is zero, when there is only one. For k_x, k_y this is without importance since for very large L we may regard k_x, k_y as continuous variables. Thus we find

$$\tfrac{1}{2}\Sigma\hbar\omega=\hbar c\frac{L^2}{\pi^2}\int\limits_0^\infty\int\limits_0^\infty\left[\tfrac{1}{2}\sqrt{k_x^2+k_y^2}+\sum_{n=1}^{\infty}\sqrt{n^2\frac{\pi^2}{a^2}+k_x^2+k_y^2}\right]dk_x\,dk_y$$

or, introducing polar coordinates in the $k_x\,k_y$ plane,

$$\tfrac{1}{2}\Sigma\hbar\omega=\hbar c\frac{L^2}{\pi^2}\cdot\frac{\pi}{2}\sum_{(0)1}^{\infty}\int\limits_0^\infty\sqrt{\left(n^2\frac{\pi^2}{a^2}+\varkappa^2\right)}\,\varkappa\,d\varkappa,$$

where the notation $(0)1$ is meant to indicate that the term with $n=0$ has to be multiplied by $\frac{1}{2}$. For very large a also this last summation may be replaced by an integral and it is therefore easily seen that our interaction energy is given by

$$\delta E=\hbar c\frac{L^2}{\pi^2}\cdot\frac{\pi}{2}\left\{\sum_{(0)1}^{\infty}\int\limits_0^\infty\sqrt{\left(n^2\frac{\pi^2}{a^2}+\varkappa^2\right)}\,\varkappa\,d\varkappa-\int\limits_0^\infty\int\limits_0^\infty\sqrt{(k_z^2+\varkappa^2)}\,\varkappa\,d\varkappa\left(\frac{a}{\pi}dk_z\right)\right\}.$$

In order to obtain a finite result it is necessary to multiply the integrands by a function $f(k/k_m)$ which is unity for $k\ll k_m$ but tends to zero sufficiently rapidly for $(k/k_m)\to\infty$, where k_m may be defined by $f(1)=\frac{1}{2}$. The physical meaning is obvious: for very short waves (X-rays e.g.) our plate is hardly an obstacle at all and therefore the zero point energy of these waves will not be influenced by the position of this plate.

Introducing the variable $u=a^2\varkappa^2/\pi^2$,

$$\delta E=L^2\hbar c\frac{\pi^2}{4a^3}\left\{\sum_{(0)1}^{\infty}\int\limits_0^\infty\sqrt{n^2+u}\,f(\pi\sqrt{n^2+u}/a k_m)\,du\right.$$
$$\left.-\int\limits_0^\infty\int\limits_0^\infty\sqrt{n^2+u}\,f(\pi\sqrt{n^2+u}/a k_m)\,du\,dn\right\}.$$

We apply the EULER–MACLAURIN formula:

$$\sum_{(0)1}^{\infty}F(n)-\int\limits_0^\infty F(n)\,dn=-\tfrac{1}{12}F'(0)+\tfrac{1}{24\times 30}F'''(0)+\dots$$

Introducing $w=u+n^2$ we have

$$F(n)=\int\limits_{n^2}^{\infty}w^{1/2}f(w\pi/a k_m)\,dw,$$

---

[1] H. B. G. CASIMIR and D. POLDER, Phys. Rev., 73, 360 (1948).

whence

$$F'(n) = -2n^2 f(n^2 \pi / a k_m)$$
$$F'(0) = 0$$
$$F'''(0) = -4.$$

The higher derivatives will contain powers of $(\pi / a k_m)$. Thus we find

$$\delta E / L^2 = -\hbar c \frac{\pi^2}{24 \times 30} \cdot \frac{1}{a^3},$$

a formula which holds as long as $a k_m \gg 1$. For the force per cm^2 we find

$$F = \hbar c \frac{\pi^2}{240} \frac{1}{a^4} = 0{,}013 \frac{1}{a_\mu^4} \text{ dyne/cm}^2$$

where a_μ is the distance measured in microns.

We are thus led to the following conclusions. There exists an attractive force between two metal plates which is independent of the material of the plates as long as the distance is so large that for wave lengths comparable with that distance the penetration depth is small compared with the distance. This force may be interpreted as a zero point pressure of electromagnetic waves.

Although the effect is small, an experimental confirmation seems not unfeasable and might be of a certain interest.

Natuurkundig Laboratorium der N.V. Philips'
Gloeilampenfabrieken, Eindhoven.)

Quantum dynamics of Kaluza-Klein theories

Thomas Appelquist and Alan Chodos
J. W. Gibbs Laboratory, Yale University, New Haven, Connecticut 06511
(Received 8 April 1983)

Some of the quantum properties of Kaluza-Klein theories are studied. The classical features of these theories are reviewed, and the quantization of the gravitational field in an arbitrary number of dimensions is described. These results are then applied to a detailed analysis of the five-dimensional Kaluza-Klein model. The fifth dimension is taken to be compact and a quantum effective potential, as a function of the five-five component of the metric, is constructed. It is argued that the one-loop computation is reliable as long as the distance around the fifth dimension is large compared to the Planck length. The effective potential separates into two pieces: an induced cosmological constant, independent of the size of the fifth dimension, and a distance-dependent "Casimir" energy. The cosmological term is subtracted, leaving an attractive Casimir potential which will contract the fifth dimension to a size on the order of the Planck length. Consequences of this result are discussed and some of the ways in which it can be generalized are outlined.

I. INTRODUCTION

One of the most intriguing and elegant ways of unifying gauge field theories with gravitation is also one of the oldest. The suggestion of Kaluza[1] and Klein[2] was that electromagnetism and general relativity could be unified by starting with a five-dimensional version of the latter and then somehow arranging for the fifth dimension to become unobservable. In recent years, the Kaluza-Klein idea has been generalized to higher dimensions in an effort to unify non-Abelian gauge fields with gravitation.[3,4] The possibility of obtaining a realistic four-dimensional theory by starting with a simpler theory in a higher-dimensional space has been actively pursued in the context of supergravity. The construction of the $N=8$ theory by starting with $N=1$ supergravity in 11 dimensions[5] is especially attractive.

We shall take the point of view that any implementation of the Kaluza-Klein idea should regard the extra dimensions as actually existing with some physical size, rather than as only an intermediate device for deriving interesting four-dimensional theories. If so, then these extra dimensions presumably form some compact manifold whose size is extremely small, perhaps not much larger than the Planck length $(\hbar G/c)^{1/2}=1.6\times10^{-33}$ cm.

In a recent letter,[6] we suggested that the vacuum fluctuations of the higher-dimensional gravitational field might provide a physical mechanism capable of accounting for the extreme smallness of the extra dimensions. An explicit computation was carried out in the five-dimensional prototypical Kaluza-Klein model which showed that these quantum fluctuations give rise to a gravitational Casimir effect[7] which can contract the fifth dimension to a size on the order of the Planck length. It was assumed that the relevant vacuum solution to the field equation of five-dimensional general relativity is the Cartesian product of flat four-space and a compact fifth dimension with associated metric component g_{55}. The vacuum fluctuations about this classical background are then computed in the form of a quantum effective potential as a function of g_{55}. When the distance around the fifth dimension is larger than the Planck length, the loop expansion is reliable and the potential is seen to give rise to an attractive Casimir force.

The purpose of this paper is to describe the quantum physics of the five-dimensional model in more detail and to begin to discuss some of its generalizations. The most important of these is the extension to higher numbers of dimensions in order to accommodate non-Abelian gauge fields. In this paper, our remarks along these lines will be restricted to the classical theory. Some of these have appeared already in the literature[3,4] but we include them here in order to prepare for a study of the quantum properties of these theories.

The effective dimensional reduction which takes place in a Kaluza-Klein theory is analogous to the effective reduction of a high-temperature four-dimensional gauge theory to three dimensions.[8] Such theories have a periodicity in the time coordinate inversely proportional to the temperature. For temperatures much larger than the momentum scales being considered, the time dimension is squeezed out of the problem, leaving an effective massless three-dimensional theory. This classical picture is then modified by quantum loops composed of the nonzero (massive) modes in the Fourier time expansion. Because of the ultraviolet singularities in certain loop computations, they do not completely decouple from the low-momentum physics. In particular, they produce an effective mass (a Debye mass) for the electric components of the gauge field. Thus quantum effects screen the electric field at large distances, completing the reduction to an effective three-dimensional theory. The effective potential computation in Kaluza-Klein theories is very similar to the Debye-screening computation in finite-temperature gauge theories.

In Sec. II, we discuss the classical features of Kaluza-Klein theories. The five-dimensional model and its reduction to four dimensions is described in some detail. We relate the time-independent solution, which will be used as a starting point for the quantum computations, to a time-dependent classical solution (the Kasner solution) that was considered in Ref. 9. Some of the features of higher-dimensional compactification are then described. Finally, returning to five dimensions, we show that the presence of a cosmological term which is naturally generated by the quantum fluctuations, leads to a five-dimensional de Sitter-type solution. With the assumption that the cosmological constant is sufficiently small, this solution reduces to the above time-independent or Kasner solutions for arbitrarily long times.

In Sec. III, some of the properties of quantum gravity in D dimensions are described. The quantization of the five-dimensional theory is carried out in detail, using a particularly useful and physical gauge choice. The quadratic Lagrangian sufficient to carry out a one-loop computation of the effective potential is constructed, the gauge-fixing and ghost terms are derived, and a convenient method of regularization is described.

In Sec. IV, we present the explicit details of the computation of the one-loop effective potential as a function of g_{55}. In Sec. V, this result is interpreted, the analogy with the Casimir effect is explored, and the validity of the loop expansion is discussed. Finally, for the convenience of the reader, we include an Appendix in which some of the relevant features of finite-temperature gauge theories are summarized.

II. CLASSICAL FEATURES OF KALUZA-KLEIN THEORIES

To begin, we focus on the classical theory of general relativity, not in the usual 3 + 1 dimensions but instead in $D=d+1$ dimensions. The action is

$$S=\frac{1}{16\pi G_D}\int d^Dx\sqrt{\bar{g}}R\ , \tag{2.1}$$

where $\bar{g}_{AB}$ is the metric in D-dimensional space, $1\leq A,B\leq D$, R is the scalar curvature computed from $\bar{g}$, and G_D is the D-dimensional version of Newton's constant. It will be convenient for our purposes to work with the Euclidean version of the theory in which the signature of $\bar{g}_{AB}$ is D. Our units are such that $\bar{g}_{AB}$ is dimensionless; since R contains two derivatives and since S must be dimensionless (we have set $\hbar=1$), it follows that the dimension of G_D is (length)$^{D-2}$. For the time being, we omit a cosmological constant in the action, Eq. (2.1). The possible role of such a constant will be discussed later.

The central idea of the Kaluza-Klein approach is that "dimensional reduction" takes place; i.e., the D-dimensional theory becomes effectively equivalent to a four-dimensional theory, with the extra $D-4$ dimensions being somehow rendered unobservable. In the literature one finds two points of view regarding dimensional reduction. On the one hand, some authors use it as a technical device for obtaining rather complicated Lagrangians in four dimensions from simpler Lagrangians in D dimensions. According to this approach, one merely truncates the D-dimensional theory by assuming, without need for further justification, that either the fields are totally independent of the extra $D-4$ coordinates, or else that they depend on them in some particularly simple way that can be integrated out.[10]

It is tempting, however, to suppose that the extra dimensions really exist even though we cannot detect them. This is possible if they form a compact manifold whose size is very small, perhaps as small as the Planck length $(G_D)^{1/(D-2)}$, the only length scale appearing in the action, Eq. (2.1). There is, of course, nothing in the classical theory which would force the compact dimensions to be related to the Planck length. In the classical theory, G_D is simply an overall factor and is irrelevant. In the quantum theory, we will discover a mechanism which might be responsible for the contraction of the extra dimensions to the order of the Planck length. However, there will still remain the question of why the Planck length, which is very different from all other length scales in physics, arises in the first place.

To see in more detail how dimensional reduction works, it is simplest to begin with the five-dimensional model. One chooses coordinates x^μ for (3 + 1)-dimensional space-time ($\mu=0,1,2,3$) and x^5 for the extra dimension which is assumed to be a circle of radius R_5:

$$0\leq x^5<2\pi R_5\ . \tag{2.2}$$

Here R_5 is a parameter with dimensions of length. We observe that, by itself, R_5 has no physical significance; it merely serves to characterize the range of the coordinate x_5. What can have meaning, however, is the distance around the extra dimension:

$$\delta_5=\int_0^{2\pi R_5}dx^5\sqrt{\bar{g}_{55}}\ . \tag{2.3}$$

[Even this could have no significance if the theory were truly generally covariant in a five-dimensional sense. The field equations derived from Eq. (2.1) will be generally covariant, but the boundary condition we have chosen, namely, that the extra dimension be compact, breaks the five-dimensional general covariance.]

To proceed, we parametrize the metric in the form

$$\bar{g}_{AB}=\phi^{-1/3}\begin{bmatrix} g_{\mu\nu}+A_\mu A_\nu\phi & A_\mu\phi \\ A_\mu\phi & \phi \end{bmatrix} \tag{2.4}$$

and we expand each component in a Fourier series

$$\bar{g}_{AB}(x^\mu,x^5)=\sum_{n=-\infty}^{\infty}\bar{g}_{AB}^{(n)}(x^\mu)e^{inx^5/R_5} \tag{2.5}$$

So far, no approximations have been made. Any five-dimensional metric can be written in the form of Eq. (2.4) and any sufficiently smooth function on a finite interval can be expanded in the form of Eq. (2.5). Complete dimensional reduction is then achieved by keeping only the $n=0$ (x^5-independent) mode of $\bar{g}_{AB}$. Upon so doing, one finds

$$S_4=-\frac{1}{16\pi G}\int d^4x(\det\{g_{\mu\nu}\})^{1/2}\left[R^{(4)}+\tfrac{1}{4}\phi F_{\mu\nu}F^{\mu\nu}+\frac{1}{6}\frac{\partial_\mu\phi\partial^\mu\phi}{\phi^2}\right]. \tag{2.6}$$

Here $R^{(4)}$ is the four-dimensional scalar curvature computed from $g_{\mu\nu}$, $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$, and $G=G_5/2\pi R_5$. (The zero-mode superscript on the fields has been dropped.) The role of the Weyl factor $\phi^{-1/3}$ in Eq. (2.4) is to ensure that no extra factors of ϕ multiply $R^{(4)}$ in Eq. (2.6). Thus the first term in S_4 looks precisely like general relativity. This choice is purely cosmetic, however; Eq. (2.6) really represents a scalar-tensor theory of gravitation, and it is not possible to isolate gravitational effects in $g_{\mu\nu}$ alone or in ϕ alone. Indeed, with $A_\mu=0$, the action S_4 describes a particular version of the well-known Brans-Dicke theory[11] (the Brans-Dicke parameter ω is zero in this case). Had we chosen some other Weyl factor, S_4 would have looked different in detail, but all physical predictions would be the same.

Equation (2.6) also illustrates the simplest instance of the Kaluza-Klein "miracle": the appearance of the standard gauge piece $F_{\mu\nu}F^{\mu\nu}$. It is true that in order to make it take the conventional form, one had to be clever in one's choice of parametrization, Eq. (2.4), but it is not true that the gauge fields were put in by hand; the starting point was pure general relativity in five dimensions.

In order to provide a physical mechanism for dimensional reduction, it will be important to work within the full five-dimensional theory [all terms in the Fourier series Eq. (2.5) must, *a priori*, be considered]. Quantum corrections will be computed as an expansion in $\hbar$ about a "vacuum" classical solution to the five-dimensional field equations derived from the action Eq. (2.1). Using the nomenclature of Eq. (2.4), the classical solution of choice is

$$\begin{aligned}&g_{\mu\nu}=\delta_{\mu\nu}\,,\\&A_\mu=0\,,\\&\phi=\phi_c=\text{constant}\,,\end{aligned} \tag{2.7}$$

along with the manifold restriction Eq. (2.2). The distance around the fifth dimension, Eq. (2.3), will be determined by the magnitude of ϕ_c which, in turn, will depend on the quantum corrections. The question of why the five-dimensional theory chooses the classical solution, Eq. (2.7), over some other remains unanswered. Nevertheless, the seemingly trivial fact that it *is* a solution is worth emphasizing since corresponding solutions do not always exist in more realistic, higher-dimensional theories.

The classical solution, Eq. (2.7), is not the only one which might be used as a starting point for the five-dimensional Kaluza-Klein theory. In Ref. 9, it was pointed out that there exists a very simple solution to the five-dimensional Einstein equations (the Kasner solution), in which the metric has a time dependence of an especially appropriate form. The manifold is one in which time ranges over the entire real line and all the spatial coordinates x_i ($i=1,2,3$), x_5 have a finite range, say $0\le x_i, x_5<R$. The metric has the form

$$\begin{aligned}&\phi^{-1/3}g_{00}=1,\ \phi^{-1/3}g_{ij}=(t/t_0)^{2p_1}\delta_{ij}\,,\\&A_\mu=0\,,\\&\phi^{2/3}=(t/t_0)^{2p_2}\,,\end{aligned} \tag{2.8}$$

where the sum rules

$$\begin{aligned}&3p_1+p_2=1\,,\\&3{p_1}^2+{p_2}^2=1\end{aligned} \tag{2.9}$$

must be obeyed. The interesting case is $p_1=\frac{1}{2}$, $p_2=-\frac{1}{2}$ in which the fifth dimension shrinks and the three other spatial dimensions expand with time. (The case $p_1=0$, $p_2=1$ is simply flat space in disguise.)

The evolution of the metric over cosmological time scales might then explain both the smallness of the fifth dimension and the "largeness" of the visible universe. The time-independent solution, Eq. (2.7), with ϕ_c small, can then be regarded as an approximation to the Kasner solution during the present epoch. We shall see in Sec. IV that this approximation will be adequate for the computation of the quantum effective potential. It will be the quantum effects which finally determine the size of the fifth dimension relative to the Planck length.

We next consider how the classical theory generalizes when $D>5$ in order to expose some of the difficulties which are not apparent from the five-dimensional case and which are often not pointed out in the literature. Nevertheless, since the quantum corrections will be computed only for $D=5$ in the following sections, the reader who is specifically interested in those computations may wish to rejoin the discussion in the paragraph following Eq. (2.20).

For the case $D>5$, it is natural to write

$$\bar{g}_{AB}=W(\phi)\begin{bmatrix} g_{\mu\nu}+B^a_\mu B^b_\nu\phi_{ab} & B^c_\mu\phi_{ca}\\ B^c_\nu\phi_{cb} & \phi_{ab}\end{bmatrix}, \tag{2.10a}$$

with inverse

$$\bar{g}^{AB}=W^{-1}(\phi)\begin{bmatrix} g^{\mu\nu} & -B^a_\lambda g^{\lambda\mu}\\ -B^b_\lambda g^{\lambda\nu} & \phi^{ab}+B^a_\lambda B^b_\sigma g^{\lambda\sigma}\end{bmatrix}. \tag{2.10b}$$

Here $W(\phi)$ is an appropriately chosen Weyl factor, and one expects that $g_{\mu\nu}$ $(\mu,\nu=0,1,2,3)$ will play the role of the metric of spacetime, ϕ_{ab} $(a,b=1,2,\cdots,D-4)$ will be scalar fields, and B_{μ}^{a} will somehow represent the gauge degrees of freedom. Notice that both $\bar{g}_{AB}$ and $\bar{g}^{AB}$ are polynomial in B_{μ}^{a}, and that

$$\sqrt{\bar{g}}=\sqrt{g}\,\sqrt{\det\phi}[W(\phi)]^{D/2}\ . \tag{2.11}$$

Thus the Lagrangian will be polynomial in the gauge fields. (In fact, no higher than quartic terms will occur.)

If one computes the action, Eq. (2.1), based on the parametrization Eq. (2.8), one finds a term

$$\left[-\frac{1}{16\pi G_D}\right]\sqrt{\det\phi}[W(\phi)]^{D/2-1}$$
$$\times\sqrt{g}\left[\tfrac{1}{4}F_{\mu\nu}^{a}F_{k\lambda}^{b}g^{\mu k}g^{\nu\lambda}\phi_{ab}\right]. \tag{2.12}$$

[The natural choice for W is $W(\phi)=(\det\phi)^{-1/(D-2)}$; cf. Eq. (2.4).] Here

$$F_{\mu\nu}^{a}=B_{\mu;\nu}^{a}-B_{\nu;\mu}^{a}\ , \tag{2.13}$$

where

$$B_{\mu;\nu}^{a}=B_{\mu,\nu}^{a}-B_{\nu}^{b}B_{\mu,b}^{a}\ . \tag{2.14}$$

This is almost the generalization of the five-dimensional result to the non-Abelian case. We must somehow convert the extra terms in $F_{\mu\nu}^{a}$ into the standard $f^{abc}A_{\mu}^{b}A_{\nu}^{c}$ form. The way to do this is to assume the existence, on the internal manifold, of a set of vector fields ξ_{α}^{a} which obey the relations

$$[\zeta_{\alpha},\zeta_{\beta}]^{a}\equiv\zeta_{\alpha}^{b}\zeta_{\beta,b}^{a}-\zeta_{\beta}^{b}\zeta_{\alpha,b}^{a}=-f^{\alpha\beta\gamma}\zeta_{\gamma}^{a}\ . \tag{2.15}$$

Here $[\zeta_{\alpha},\zeta_{\beta}]$ is the standard Lie bracket. We also assume that $\zeta_{\alpha,\mu}^{a}=0$. Setting

$$B_{\mu}^{a}=\zeta_{\alpha}^{a}A_{\mu}^{\alpha} \tag{2.16}$$

it then follows that

$$F_{\mu\nu}^{a}=\zeta_{\alpha}^{a}(A_{\mu,\nu}^{\alpha}-A_{\nu,\mu}^{\alpha})-A_{\nu}^{\beta}\zeta_{\beta}^{b}(\zeta_{\alpha}^{a}A_{\mu}^{\alpha})_{,b}+A_{\mu}^{\beta}\zeta_{\beta}^{b}(\zeta_{\alpha}^{a}A_{\nu}^{\alpha})$$
$$=\zeta_{\alpha}^{a}(A_{\mu,\nu}^{\alpha}-A_{\nu,\mu}^{\alpha}-f^{\alpha\beta\gamma}A_{\mu}^{\beta}A_{\nu}^{\gamma})+\cdots\ , \tag{2.17}$$

where the ... refers to terms that depend on the derivatives of A_{μ}^{α} with respect to the internal coordinates.

A problem with the non-Abelian generalization of the Kaluza-Klein idea is that, whereas in the five-dimensional case the manifold $M^4\times S^1$, with metric given in Eq. (2.7) solves the Einstein equations and hence serves as an appropriate background for the quantum dynamics, no such manifold exists in the non-Abelian case. One can see this as follows. Let us assume that the manifold we are searching for has a "vacuum" metric of the form

$$\bar{g}_{AB}=W(\phi)\begin{bmatrix}\delta_{\mu\nu} & 0\\ 0 & \phi_{ab}(y)\end{bmatrix}. \tag{2.18}$$

This is the analog of the vacuum solution, Eq. (2.7), in the Abelian case. Here we use y to denote the coordinates of the internal manifold. Inserting this ansatz into the Einstein equations, one finds, first, that W must be a constant and, second, that

$$R_{ab}(\phi)=0\ . \tag{2.19}$$

$R_{ab}(\phi)$ is the Ricci tensor computed from ϕ_{ab} alone. The solution to Eq. (2.19) is complicated by the fact that we need the vector fields introduced in Eq. (2.15). Indeed, for the vacuum solution, one would like these vectors to be symmetries of the metric ϕ_{ab}:

$$\nabla_a\zeta_{ab}+\nabla_b\zeta_{aa}=0\ , \tag{2.20}$$

where the covariant derivative ∇_a is with respect to the metric ϕ. However, there is a theorem[12] which states that Eqs. (2.15) (with nonvanishing f's), (2.19) and (2.20) are incompatible on a compact Riemannian manifold. If one gives up (2.15), one gives up non-Abelian gauge theories; if one abandons Eq. (2.19), one abandons Einstein's equation; and if one forsakes Eq. (2.20), one has a hard time understanding what it means to have a natural candidate for the vacuum field configuration. In the literature, the usual assumption is that the pure Einstein equations no longer apply, either because of the addition of extra matter fields,[13] or because of the presence of torsion,[14] or because of quantum effects.[15] The danger to be avoided, however, is that in a more complicated theory the assumption of extra dimensions may fail to explain anything, such as the origin of gauge theories. Under such circumstances, one might as well stick to four dimensions.

To close this section, we return briefly to the question of a cosmological constant. We shall see in Sec. IV that, just as in any nonsupersymmetric field theory, the fluctuations of the quantum fields induce a constant, infinite, vacuum energy density, which in a theory of gravity plays the role of a cosmological constant. In the usual spirit of renormalization, it becomes unnatural to expect that the observed cosmological constant should be zero; in fact, its natural size exceeds the observed upper bound by 120 orders of magnitude.

There is no commonly accepted resolution of this inconsistency. For the purposes of this paper, we shall simply take the cosmological constant to vanish as an observational fact, and we shall tune our parameters at each stage of the calculation to ensure that it remains zero, however unnatural an act that may be.

To see the dynamical role of a cosmological term, we look at the five-dimensional Einstein equations in the presence of a cosmological constant Λ. In four dimensions, one has the usual de Sitter and anti-de Sitter solutions, depending on the sign of Λ. In the five-dimensional case, we postulate the form

$$ds^2=-dt^2+f(t)d\sigma_k{}^2+g(t)(dx^5)^2\ . \tag{2.21}$$

Here $d\sigma_k{}^2$ represents the line element of a maximally symmetric three-dimensional space; there are three such spaces, of positive $(k=1)$, zero $(k=0)$, or negative $(k=-1)$ constant curvature. We solve the Einstein equations

$$R_{AB}-\tfrac{1}{2}\bar{g}_{AB}R+\Lambda\bar{g}_{AB}=0\ , \tag{2.22}$$

or equivalently

$$R_{AB}=\alpha\bar{g}_{AB}\ , \tag{2.23}$$

$$\alpha=\tfrac{2}{3}\Lambda\ . \tag{2.24}$$

It is convenient to set $\phi=\dot{f}/f$; $\gamma=\dot{g}/g$. Equations (2.23) reduce to

$$\alpha=\tfrac{3}{2}\dot{\phi}+\tfrac{3}{4}\phi^2+\tfrac{1}{2}\dot{\gamma}+\tfrac{1}{4}\gamma^2\ , \tag{2.25}$$

$$\alpha=\frac{k}{3f}+\tfrac{1}{2}\dot{\phi}+\tfrac{1}{4}\phi(3\phi+\gamma)\ , \tag{2.26}$$

$$\alpha=\tfrac{1}{2}\dot{\gamma}+\tfrac{1}{4}\gamma(3\phi+\gamma)\ . \tag{2.27}$$

Together, Eqs. (2.25) and (2.27) imply that

$$\gamma=\frac{2\dot{\phi}}{\phi}+\phi\ , \tag{2.28}$$

i.e., that

$$g=g_0\frac{\dot{f}^2}{f}\ . \tag{2.29}$$

Then Eq. (2.26) determines f to be

$$f(t)=Ae^{\omega t}+Be^{-\omega t}+\frac{k}{3\alpha}\ , \tag{2.30}$$

where $\omega^2=\alpha$ and A and B are constants of integration. If $\alpha>0$, A and B are each real and arbitrary; if $\alpha<0$, $B=A^*$ with A arbitrary.

Of course, if Λ is small enough, we can, by setting $k=0$ in Eq. (2.30), return arbitrarily closely to the time-independent solution, Eq. (2.7). Here we show, in addition, that in the same limit one can also recover the Kasner solution, Eq. (2.8).

To achieve this, we look at the case $k=0$, $\alpha>0$, and A,B of opposite sign. Then by choosing the origin of time so that $f(0)=0$, we have the solution

$$f(t)=A_0\sinh\omega t\ , \tag{2.31}$$

$$g(t)=(A_0g_0\alpha)\left[\frac{\cosh^2\omega t}{\sinh\omega t}\right]\ . \tag{2.32}$$

To get a nontrivial limit as $\alpha\rightarrow 0$, it is necessary to take $A_0=B_0/\omega$ where B_0 remains finite. Then, provided $\omega t\ll 1$, we find

$$f(t)\sim B_0t\ , \tag{2.33}$$

$$g(t)\sim\frac{B_0g_0}{t}\ . \tag{2.34}$$

This is the Kasner-type behavior that we sought. Notice however, that when $\omega t\gg 1$, we have

$$f(t)\sim A_0e^{\omega t}\ , \tag{2.35}$$

$$g(t)\sim(A_0g_0\alpha)e^{\omega t} \tag{2.36}$$

Thus, if α is sufficiently small, there can be an arbitrarily long Kasner-type epoch; nevertheless, unless α is strictly zero, the fifth dimension will begin to grow (the changeover from contraction to expansion occurs when $e^{\omega t}=\sqrt{2}+1$), and will ultimately reemerge from the obscurity of the submicroscopic world.

III. QUANTUM GRAVITY IN *D* DIMENSIONS

A. General remarks

Classical gravitational theories, like all classical dynamical systems, must become inadequate in regimes where quantum effects can no longer be neglected. In the case of gravity, this happens when distance scales on the order of the Planck length are being approached. When the Planck length is reached, however, the effective gravitational self-coupling will become of order 1 and the loop expansion will break down. Thus, unless important quantum effects can be identified at distance scales larger than the Planck length, the conventional loop-expansion approach to the quantization of the gravitational field may not be useful.

Kaluza-Klein theories offer just such a distance scale. The radius of curvature of the extra dimension(s) can, *a priori*, be larger than the Planck length and, as we shall show in the next section, it can provide the scale for an important and reliably computable quantum effect. In order to set the stage for this computation, we briefly review some of the main features of the quantization of general relativity. We emphasize that, even if Einstein's theory is only an effective theory for use at large distances beyond the Planck length, the kind of situation which arises in Kaluza-Klein theories could make the quantization program meaningful and useful.

The path integral quantization of a field theory proceeds by first finding an extremum of the action—a classical solution, as discussed in the previous section—and then computing the quantum corrections as an expansion in $\hbar$. A natural way of implementing this program, especially if general properties such as renormalizability are being studied, is the background-field method. The fields are split into a classical part and a quantum part and the action is then expanded in the quantum fields about arbitrary classical background fields. An expansion to second order is sufficient to generate all one-loop diagrams with any number of external lines attached. Physics is then extracted by putting the external lines on the mass shell, that is, by taking the background fields to satisfy the classical field equations.

In general relativity, the background-field method[16] has been extensively used to construct the on-shell counterterms for the study of renormalizability.[17,18] The result of these studies, as expected from the fact that Newton's constant is dimensionful, is that the theories are nonrenormalizable. This means that not only is one restricted to using them to construct low energy expansions at distances beyond the Planck length, but the coefficients in this expansion must be taken to be arbitrary parameters. We expect this to be the case in higher-dimensional theories as well but we will see that some quantities might be reliably computed at distances beyond the Planck length. In particular, quantum effective potentials which govern the dynamics of the extra compact dimensions can be computed when these dimensions are larger than the Planck length.

B. Quantizing the five-dimensional theory

We shall consider the pure Einstein theory, Eq. (2.1), in five dimensions, taking as a starting classical solution the generalized Minkowski space, Eq. (2.7), with $0 \le x^5 < 2\pi R_5$. The effect of an additive cosmological term, which is naturally induced by the quantum corrections, was discussed in Sec. II and we shall briefly return to it at the end of this section.

The metric $\bar{g}_{AB}$ is parametrized as in Eq. (2.4) and quantum fields $h_{\mu\nu}$, A_μ, and ϕ' are then introduced by setting

$$\begin{aligned} g_{\mu\nu} &= \delta_{\mu\nu} + h_{\mu\nu} , \\ \phi &= \phi_c + \phi' . \end{aligned} \tag{3.1}$$

The action, Eq. (2.1), can then be expanded as an infinite series in the quantum fields. Note that even though these fields have been introduced with an eye toward dimensional reduction, they are at this point still functions of x_μ and x_5.

To determine propagators and to compute one-loop diagrams in the classical background field, it suffices to keep only terms quadratic in the quantum fields. The result of this exercise is

$$S^{(1)} = -\frac{1}{16\pi G_5} \int d^5x \, \mathscr{L}^{(2)} , \tag{3.2}$$

where

$$\mathscr{L}^{(2)} = \mathscr{L}_{\text{Gr}} + \mathscr{L}_G + \mathscr{L}_S + \mathscr{L}_5 \tag{3.3}$$

and

$$\begin{aligned} \mathscr{L}_{\text{Gr}} = \tfrac{1}{4} (& h_{\mu\nu,\rho} h_{\mu\nu,\rho} - 2 h_{\mu\nu,\rho} h_{\mu\rho,\nu} \\ & - 2 h_{\mu\nu,\mu} h_{\rho\rho,\nu} + h_{\rho\rho,\mu} h_{\sigma\sigma,\mu}) \end{aligned}$$

$$+ \frac{1}{4\phi_c} (h_{\mu\nu,5} h_{\mu\nu,5} - h_{\mu\mu,5} h_{\nu\nu,5}) , \tag{3.4}$$

$$\mathscr{L}_G = \tfrac{1}{4} \phi_c (F_{\mu\nu})^2 , \tag{3.5}$$

$$\mathscr{L}_S = \frac{1}{6{\phi_c}^2} (\partial_\mu \phi')^2 , \tag{3.6}$$

$$\begin{aligned} \mathscr{L}_5 = & -\frac{3}{4\phi_c} \left[\left[\frac{\phi'_{,5}}{\phi_c} \right]^2 - h_{\mu\mu,5} \frac{\phi'_{,5}}{\phi_c} \right] \\ & - \frac{3}{2} \frac{\phi'_{,5}}{\phi_c} \partial_\mu A_\mu + A_{\mu,5} (h_{\nu\nu,\mu} - h_{\mu\nu,\nu}) . \end{aligned} \tag{3.7}$$

There are several points worth making about this result:

(1) There is no mixing among $h_{\mu\nu}$, A_μ, and ϕ' in the $n=0$ sector. This was, of course, already demonstrated in the full zero-mode Lagrangian, Eq. (2.6), and is a consequence of the variable choice, Eq. (2.4), in particular the presence of the Weyl factor $\phi^{-1/3}$.

(2) The first piece $\mathscr{L}_{\text{Gr}}$, Eq. (3.4), is the usual quadratic Lagrangian for the graviton, together with a masslike term for the $n \neq 0$ modes. It can be inverted for $n \neq 0$ to give the unitary form of the massive, spin-2 propagator.

(3) The second two pieces, $\mathscr{L}_G$ and $\mathscr{L}_S$, describe the massless gauge field and scalar field.

(4) The role of the constant, classical background field ϕ_c as the metric for the fifth dimension can be seen in each term.

A simple and natural gauge choice for the quantization of this theory is the cylindrical gauge

$$\bar{g}_{\mu 5,5}(x) = 0 . \tag{3.8}$$

It is analogous to the static gauge in finite temperature gauge theories, discussed in the Appendix. It can be effected by the addition of the gauge-fixing Lagrangian

$$\Delta \mathscr{L}_{\text{gf}} = -\tfrac{1}{2} (\Lambda^B \bar{g}_{AB})^2 , \tag{3.9}$$

where

$$\Lambda^B = (\Lambda^\mu, \zeta \partial_5) , \tag{3.10}$$

and Λ^μ is, for the moment, an arbitrary four vector. The cylindrical gauge, Eq. (3.8), is then obtained by taking the limit $\zeta \to \infty$. Only the $n=0$ mode in the Fourier expansion, Eq. (2.5), will remain sensitive to Λ^μ in this limit. Using the variables $h_{\mu\nu}$, A_μ, and ϕ', and referring to the quadratic Lagrangian, Eqs. (3.3)–(3.7), Λ^μ can be used for further gauge fixing within the massless, zero-mode sector.

In the cylindrical gauge, the fields A_μ and ϕ are independent of x_5—they are purely zero mode. Because of this, the $\mathscr{L}_5$ term in the quadratic Lagrangian, Eqs. (3.3) and (3.7), which mixes the fields, must vanish. The underlying reason for this useful simplification is that the cylindrical gauge is *the* physical gauge for the five-dimensional Kaluza-Klein theory. To see this, we note that in N dimensions, an $N\times N$ metric field g_{AB} describes $(N^2+N)/2-2N$ physical degrees of freedom, the subtraction of $2N$ corresponding to the freedom in the choice of a coordinate basis and the fact that g_{A0} are dependent variables. The 5 degrees of freedom in the five-dimensional theory are then accounted for in the following way.[19] In the zero-mode sector, they correspond to the massless graviton, photon, and Brans-Dicke scalar. By contrast, the 5 degrees of freedom in each nonzero mode are those of a single massive, spin-2 particle. Clearly the cylindrical gauge explicitly reflects the physics of the five-dimensional Kaluza-Klein theory.

The absorption of spin-0 and spin-1 degrees of freedom to create a massive spin-2 particle is, of course, a Higgs mechanism. The chosen classical solution [Eqs. (2.2) and (2.7)] involving, as it does, a fixed scale associated with the fifth dimension, clearly does not have all the symmetries of the original Lagrangian. Spontaneous symmetry breakdown ("spontaneous compactification") has taken place for those modes ($n\neq 0$) which are sensitive to the presence of the scale R_5, and the fields A_μ and ϕ are the associated Goldstone fields.

The change in $\bar{g}_{AB}$ induced by an infinitesimal coordinate transformation

$$x^A\rightarrow x^A-\epsilon^A(x) \tag{3.11}$$

is

$$\delta\bar{g}_{AB}(x)=\epsilon^C\bar{g}_{AB,C}+\epsilon^C{}_{,A}\bar{g}_{CB}+\epsilon^C{}_{,B}\bar{g}_{AC}\ . \tag{3.12}$$

From this, one finds in the usual way that the DeWitt-Faddeev-Popov ghost factor Δ is

$$\Delta=\det M\ , \tag{3.13}$$

where M is defined by

$$\int M_{AB}(x,x')\epsilon^B(x')dx'=[\bar{g}_{A5,B}\partial_5+\bar{g}_{B5}\partial_A\partial_5+\bar{g}_{AB}(\partial_5)^2+\bar{g}_{AB,5}\partial_5]\epsilon^B(x)\ . \tag{3.14}$$

The gauge condition $\bar{g}_{A5,5}=0$, corresponding to the limit $\zeta\rightarrow\infty$ in Eq. (3.10), has been used to simplify this expression. The quantity Δ may be represented as an integral over anticommuting variables:

$$\Delta=\int D\eta\, D\bar{\eta}\exp\left[\int dx\, dx'\bar{\eta}^A(x)M_{AB}(x,x')\eta_B(x')\right]\ . \tag{3.15}$$

We see that the ghost action is already quadratic in the fields η and $\bar{\eta}$, and hence the one-loop approximation is obtained by setting $\bar{g}_{AB}$ in Eq. (3.14) to the background value Eq. (2.7).

Having obtained the one-loop versions of the classical action and the gauge-fixing and ghost terms, we are almost ready to study the quantum theory defined by

$$z=\int D\mu(h)\Delta(h)e^{-[S^{(2)}+S_{\rm gf}]}\ , \tag{3.16}$$

where

$$\bar{g}_{AB}=g_{AB}^{(0)}+h_{AB}$$

and $g_{AB}^{(0)}$ is given in Eq. (2.7). The only remaining issue is what to take for the measure $D\mu(h)$. This question has been discussed at length by Fradkin and Vilkovisky,[20] and by DeWitt[21] and 't Hooft,[22] among others. Its definitive resolution probably awaits a well-defined quantum theory of gravity. However, it is clear that the measure must at least intervene to cancel unwanted terms in physical quantities proportional to $\delta^4(0)$ and, in our case, $\delta^5(0)$. This role of the measure can be seen, for example, in derivatively coupled scalar theories in four dimensions, where Feynman graphs with one or more loops typically contain quartic divergences. These are precisely canceled by a factor in the measure that can be derived by starting with the phase-space path integral where the measure factor is unity and then integrating out the canonical momenta to obtain the standard path integral in configuration space.

In performing our computation, we shall enforce this cancellation by setting $D\mu(h)=Dh$, and at the same time systematically deleting any terms in the effective potential which are proportional to $\delta^4(0)$ or $\delta^5(0)$. This is consistent with having chosen a regularization scheme such as dimensional continuation or zeta-function regularization, in which these singular factors would have been defined to zero from the beginning.

C. Divergences and counterterms

In D-dimensional quantum gravity, the maximum degree of divergence of an L-loop graph is

$$d=(D-2)L+2\ . \tag{3.17}$$

This maximum, independent of the number of external lines, is attained only if the two momentum factors at each vertex are taken to be internal loop momenta. The counterterm being computed would then have no derivatives acting on the fields; it would correspond to an induced cosmological constant. In general, an L-loop amplitude will produce terms ranging from finite ones (thank goodness) up to the maximally divergent cosmological constant. The possible divergent terms will be dictated by the local counterterms allowed by D-dimensional general covariance. The actual divergences encountered in a computation will further depend on the regulation scheme employed. We here restrict ourselves to some general remarks about counterterms, imagining that a cutoff Λ is in place on the momentum integrals. A specific such cutoff will be introduced in the next section in the course of describing the effective potential computation.

At one loop, the maximal degree of divergence is $d=D$. For the five-dimensional theory, the leading quintic divergence corresponds to an induced cosmological constant. It must, of course, be fine-tuned away and we have nothing new to add to this age-old problem. A cubic divergence will appear multiplying the scalar curvature R and it can, therefore, be absorbed into a renormalization of Newton's constant. Finally, there can be a linear divergence multiplying the four-derivative operator $(R_{\mu\nu\lambda\sigma})^2$. In an even number of dimensions, this operator must vanish on shell[12] because it is proportional to $R^2-4R_{\mu\nu}{}^2$ plus a total divergence. In five dimensions, it persists and one sees the nonrenormalizability already at one loop. Of course, all these divergences would be defined to be zero using dimensional continuation or zeta-function regularization.

The effective potential is a zero-derivative operator and it is therefore a (finite) partner of the induced cosmological constant. It will be possible to separate these two pieces and then to focus our attention on the finite part. The presence of the other, higher-derivative, counterterms at one loop will not impact on the effective potential. In higher orders, all these terms can be mixed together and this problem will be briefly discussed after computing the effective potential.

IV. COMPUTATION OF THE EFFECTIVE POTENTIAL

Having outlined the general framework, we turn to the explicit evaluation of the effective potential. We begin with

$$Z[\phi_c]=\int Dh\,\delta(h_{A5,5})\Delta(h)e^{-S[h]}\ , \tag{4.1}$$

where we have expanded the metric as in Eq. (3.1). We have chosen to take the $\zeta\to\infty$ limit of the gauge-fixing Lagrangian, Eqs. (3.9) and (3.10), immediately; hence the delta functional appears in Eq. (4.1). To compute the effective potential to one loop, we use[23]

$$Z[\phi_c]=\exp[-V_{\text{eff}}(\phi_c)(g^{(0)})^{1/2}\int d^5x]\ . \tag{4.2}$$

This differs from the usual expression in that we have extracted a factor $(g^{(0)})^{1/2}$ in defining V_{eff}. If we did not do this, V_{eff} would not be generally covariant, and hence would be without physical significance.

For a one-loop computation, we need only the terms in S that are quadratic in the quantum fields. The relevant Lagrangian has been given in Eqs. (3.4)–(3.6). We can see explicitly from these equations how it is that the zero modes (the scalar ϕ', the vector A_μ, and the x_5-independent piece of the tensor $h_{\mu\nu}$) do not contribute to V_{eff}. We observe that by scaling the integration variables

$$\phi'\to\phi_c\phi'\ , \tag{4.3a}$$

$$A_\mu\to\frac{1}{\sqrt{\phi_c}}A_\mu\ , \tag{4.3b}$$

we can remove the dependence on ϕ_c completely from the zero-mode terms. The price we pay for this is that we generate a Jacobian factor in the path integral, e.g.,

$$D\phi'\to[\prod\phi_c]D\phi'=(D\phi')\exp\left[\sum_x\ln\phi_c\right]\ .$$

Using the relationship $(d^4x)\delta^4(0)=1$, we have

$$D\phi'\to\exp\left[\delta^4(0)\int d^4x\ \ln\phi_c\right](D\phi')\ . \tag{4.4}$$

This ill-defined Jacobian factor is precisely the sort of thing that must be absorbed into the measure of integration, as discussed in Sec. III. The same holds true for our rescaling of A_μ. Thus we conclude that the zero modes contribute a ϕ_c-independent factor to $Z[\phi_c]$, which will be absorbed into the overall normalization of the path integral [see Eq. (4.15) below].

Therefore, we need only concern ourselves with the massive-mode contribution to the path integral, and possibly with the ghost-determinant contribution as well. We shall first evaluate the massive mode integral, and then show that the ghost factor in fact makes no contribution.

To evaluate the massive-mode path integral, it is convenient to diagonalize the derivatives by means of Fourier transformation. Denoting by $\tilde{S}$ the contribution of these modes to the action, we have

$$\begin{aligned}\tilde{S}=\tfrac{1}{8}\sum_{n=-\infty}^{\infty}\int\frac{d^4k}{(2\pi)^4}h_{\mu\nu}^{(n)}(k)\{&[\tfrac{1}{2}(\delta^{\mu\lambda}\delta^{\nu\sigma}+\delta^{\mu\sigma}\delta^{\nu\lambda})-\delta^{\mu\nu}\delta^{\lambda\sigma}](k^2+m_n{}^2)+(\delta^{\mu\nu}k^\lambda k^\sigma+\delta^{\lambda\sigma}k^\mu k^\nu)\\&-\tfrac{1}{2}(\delta^{\mu\lambda}k^\nu k^\sigma+\delta^{\mu\sigma}k^\nu k^\lambda+\delta^{\nu\lambda}k^\mu k^\sigma+\delta^{\nu\sigma}k^\mu k^\lambda)\}h_{\lambda\sigma}^{(-n)}(-k)\ ,\end{aligned} \tag{4.5}$$

where

$$M_n{}^2=\frac{1}{\phi_c}\frac{n^2}{R_5{}^2}\ . \tag{4.6}$$

(It is easier to include the $n=0$ piece even though it will not contribute to V_{eff}.) If we abbreviate this expression by

$$\tilde{S}=\tfrac{1}{2}\sum_n\int\frac{d^4k}{(2\pi)^4}h_{\mu\nu}^{(n)}(k)\mathscr{S}^{\mu\nu\lambda\sigma}_{(n;k)}h_{\lambda\sigma}^{(-n)}(-k)\ , \tag{4.7}$$

then we have the standard result that

$$\int Dh_{\mu\nu}e^{-\tilde{S}}=(\det\mathscr{S})^{-1/2}\ . \tag{4.8}$$

We evaluate the determinant by finding the eigenvalues of $\mathscr{S}$ and taking their product. For each value of k and n, the index pair $(\mu\nu)$ takes on 10 values. Thus we can think of $\mathscr{S}$ as a 10×10 matrix labeled by k and n. (It is the fact that in momentum space, $\mathscr{S}$ is block diagonal with finite-dimensional blocks that allows us to evaluate $\det\mathscr{S}$ this way.)

Let $p^{\mu(j)}$ be vectors satisfying $p^{\mu(j)}k_\mu=0$ $(j=1,2,3)$. Then by explicit computation we find the following

$$\text{(a)}\quad 4\mathscr{S}^{\mu\nu\lambda\sigma}v^{(j)}_{\lambda\sigma}=M_n{}^2v^{\mu\nu(j)}\ ,$$

where

$$v^{(j)}_{\lambda\sigma}=k_\lambda p^{(j)}_\sigma+k_\sigma p^{(j)}_\lambda,\ \ j=1,2,3 \tag{4.9}$$

(i.e., there are three independent v's corresponding to the three independent choices for p^μ).

$$\text{(b)}\quad 4\mathscr{S}^{\mu\nu\lambda\sigma}w^{(i)}_{\lambda\sigma}=\lambda^{(i)}w^{\mu\nu(i)},\ \ i=1,2\ ,$$

where

$$\lambda^{(i)2}+2(k^2+M_n{}^2)\lambda^{(i)}-3M_n^{(4)}=0 \tag{4.10}$$

and $w^{\mu\nu(i)}$ are two linearly independent combinations of $\delta^{\mu\nu}$ and $k^\mu k^\nu$. Note that Eq. (4.10) implies that $\lambda^{(1)}\lambda^{(2)}=-3M_n{}^4$.

$$\text{(c)}\quad 4\mathscr{S}^{\mu\nu\lambda\sigma}X^{(j)}_{\lambda\sigma}=(k^2+M_n{}^2)X^{\mu\nu(j)},\ j=1,2\ ; \tag{4.11}$$

where the $X^{(j)\mu\nu}$ are appropriate linear combinations of $\delta^{\mu\nu}$, $k^\mu k^\nu/k^2$, and $p^{(j)\mu}p^{(j)\nu}/p^{(j)2}$. Although there would appear to be three $X^{(j)}$'s, corresponding to the three possible choices of $p^{\mu(j)}$, this is reduced to two because of the completeness relation

$$\delta^{\mu\nu}=\frac{k^\mu k^\nu}{k^2}+\sum_{j=1}^{3}\left[\frac{p^{\mu(j)}p^{\nu(j)}}{p^{(j)2}}\right].$$

$$\text{(d)}\quad 4\mathscr{S}^{\mu\nu\lambda\sigma}Y^{(j)}_{\lambda\sigma}=(k^2+M_n{}^2)Y^{(j)\mu\nu},\ \ j=1,2,3\ , \tag{4.12}$$

where $Y^{\mu\nu(j)}$ correspond to the three independent ways of choosing $p^{\mu(l)}$ and $p^{\mu(k)}$ in the expression

$$Y^{\mu\nu}=p^{\mu(l)}p^{\nu(k)}+p^{\nu(l)}p^{\mu(k)},\ (l\neq k)\ .$$

Thus the product of all the eigenvalues is given by

$$d(k,n)=-3[M_n{}^2(k^2+M_n{}^2)]^5\ . \tag{4.13}$$

There are five eigenvalues independent of k, corresponding to the five nonpropagating modes in the field $h_{\mu\nu}(k)$ (comprising a massive spin-1 and two spin-0 modes). There are also five eigenvectors with eigenvalue $(k^2+M_n{}^2)$; these are the five propagating components of spin 2. The fact that $d(k,0)=0$ is a signal that further gauge fixing is necessary in the zero-mode sector (i.e., one must choose coordinates in ordinary four-dimensional relativity). Since the zero modes do not contribute to V_{eff}, we need not concern ourselves with this problem.

We can now write, formally,

$$(\det 4\mathscr{S})^{-1/2}=\exp\left[-\tfrac{1}{2}\sum_n\sum_k\ln[-3M_n{}^{10}(k^2+M_n{}^2)^5]\right]. \tag{4.14}$$

Our first step is to normalize this expression by dividing by its value when $\phi_c=1$. That is, we compute the ratio

$$\frac{Z[\phi_c]}{Z[\phi_c=1]}=\exp\left\{-\tfrac{1}{2}\sum_n\sum_k\ln\left[\frac{1}{\phi_c{}^2}\,\frac{\phi_c(kR_5)^2+n^2}{(kR_5)^2+n^2}\right]^5\right\}. \tag{4.15}$$

This is just a specification of the arbitrary overall normalization of Z, and is without physical significance. It corresponds to adding an arbitrary constant to $\sqrt{g}\,V_{\text{eff}}$; by contrast, a cosmological constant, which *is* physical, is a constant term in V_{eff} itself. Next, we use the relationship

$$\left[\int d^4x\right]\frac{d^4k}{(2\pi)^4}=1 \tag{4.16}$$

to write

$$\frac{Z[\phi_c]}{Z[\phi_c=1]}=\exp\left\{-\tfrac{5}{2}\sum_{n=-\infty}^{\infty}\left[\frac{1}{2\pi R_5}\right]\int\frac{d^4k}{(2\pi)^4}\left[\ln\left[\frac{\phi_c(kR_5)^2+n^2}{(kR_5)^2+n^2}\right]-2\ln\phi_c\right]\left[\int d^5x\right]\right\}. \tag{4.17}$$

Once again, we drop the second term since it is of the form (const)$\times(\ln\phi_c)\times\delta^5(0)$, which can be absorbed by rescaling of the integration variable $h_{\mu\nu}(x)$ by an appropriate power of ϕ_c. This leaves us with

$$(g^{(0)})^{1/2}V_{\text{eff}}(\phi_c)-V_{\text{eff}}(\phi_c=1)=\frac{1}{2\pi R_5}\,\frac{5}{2}\sum_n\int\frac{d^4k}{(2\pi)^4}\ln\left[\frac{\phi_c(kR_5)^2+n^2}{(kR_5)^2+n^2}\right]. \tag{4.18}$$

$g^{(0)}$ is the determinant of the classical solution $g_{AB}{}^0$. Its value is $\phi_c{}^{-2/3}$.

Before proceeding further with our analysis of Eq. (4.18), let us return to the contribution of the ghost determinant, Eq. (3.14). Transforming to momentum space, we find from Eq. (3.14)

$$\begin{aligned} M_{AB}^{\text{one-loop}}(k,n) &= -k_5[k_A\bar{g}^{(0)}_{B5}+k_5\bar{g}^{(0)}_{AB}] \\ &= -k_5\phi_c{}^{-1/3}[\phi_c k_A\delta_{B5}+k_5\delta_{AB}+k_5\delta_{A5}\delta_{B5}(\phi_c-1)] \; . \end{aligned} \tag{4.19}$$

Here we have set $k_5=+n/R_5$. The eigenvalue equation for M reads

$$-k_5\phi_c{}^{-1/3}[\phi_c k_A v_5+k_5 v_A+k_5(\phi_c-1)\delta_{A5}v_5]=\lambda v_A \; . \tag{4.20}$$

If $v_5=0$, then this equation is satisfied with

$$\lambda=-k_5{}^2\phi_c{}^{-1/3} \; .$$

If $v_5\neq 0$ then setting $A=5$ we have

$$\lambda=-k_5\phi_c{}^{-1/3}[2\phi_c k_5] \; .$$

Thus we have four eigenvectors with eigenvalue $-k_5{}^2\phi_c{}^{-1/3}$, and one with eigenvalue $-2k_5{}^2\phi_c{}^{2/3}$. Their product is, therefore,

$$d_G=-2\left[\frac{n}{R_5}\right]^{10}\phi_c{}^{-2/3} \tag{4.21}$$

and the ratio $d_G(\phi_c)/d_G(\phi_c=1)$ is simply $\phi_c{}^{-2/3}$. Hence to one loop the ghost determinant contributes only a scaling term (const)$\times\ln\phi_c\times\delta^5(0)$ to the logarithm of $Z(\phi_c)/Z(\phi_c=1)$ and so does not affect $V_{\text{eff}}(\phi_c)$.

Returning to Eq. (4.18), we use the formula

$$\sum_{n=-\infty}^{\infty} f_n=\int_{-\infty}^{\infty}dz\, f(z)+\int_{-\infty+i\epsilon}^{\infty+i\epsilon}dz\frac{[f(z)+f(-z)]}{e^{-2\pi iz}-1} \tag{4.22}$$

to perform the mode sum. This yields

$$\begin{aligned} \phi_c{}^{-1/3}V_{\text{eff}}(\phi_c)-V_{\text{eff}}(\phi_c=1) &= \tfrac{5}{2}\int\frac{d^5k}{(2\pi)^5}\ln\left[\frac{\phi_c k^2+k_5{}^2}{k^2+k_5{}^2}\right] \\ &\quad +\frac{5}{2\pi(R_5)^5}\int\frac{d^4q}{(2\pi)^4}\int_{-\infty+i\epsilon}^{\infty+i\epsilon}dz\,\ln\left[\frac{\phi_c q^2+z^2}{q^2+z^2}\right]\frac{1}{e^{-2\pi iz}-1} \; . \end{aligned} \tag{4.23}$$

In the first term, we have defined the integration variable $k_5=z/R_5$; k^2, however, continues to denote the four-dimensional product $k_\mu k^\mu$. In the second term, we have defined the dimensionless variable $q_\mu=k_\mu R_5$.

Let us focus for the moment on the second term. By contour methods, one can easily establish the formula

$$\int_{-\infty}^{\infty}dz\, H(z)\ln\left[\frac{z^2+a^2}{z^2+b^2}\right]=2\pi\int_b^a dx\, H(ix) \tag{4.24}$$

provided $H(z)$ is analytic in the upper half plane, and that

$$H(z)\ln\left[\frac{z^2+a^2}{z^2+b^2}\right]\leq\frac{c}{|z|^{1+\epsilon}}$$

as $|z|\to\infty$ in the upper half plane. We have defined the logarithm to have its cut on the negative real axis, and have taken a,b, to be the positive square roots of a^2,b^2. In our case, $H(z)=1/(e^{-2\pi iz}-1)$ so the analyticity and boundedness properties are satisifed, and we have for the second term

$$\frac{5}{(2\pi)(R_5)^5}\int\frac{d^4q}{(2\pi)^4}\left[2\pi\int_q^{q\sqrt{\phi_c}}dx\frac{1}{e^{2\pi x}-1}\right] \; . \tag{4.25}$$

The x integration is easily done. Defining $w_\mu=2\pi q_\mu$, we have

$$\frac{5}{(2\pi R_5)^5}\left[\frac{1}{\phi_c}-1\right]\int\frac{d^4w}{(2\pi)^4}\ln(1-e^{-w}) \; , \tag{4.26}$$

i.e.,

$$\phi_c^{-1/3}V_{\text{eff}}(\phi_c)-V_{\text{eff}}(\phi_c=1)=\tfrac{5}{2}\int\frac{d^5k}{(2\pi)^5}\ln\left[\frac{\phi_c k^2+k_5^2}{k^2+k_5^2}\right]+\frac{5}{(2\pi R_5)^5}\left[\frac{1}{\phi_c^2}-1\right]\int\frac{d^4w}{(2\pi)^4}\ln(1-e^{-w})\ . \tag{4.27}$$

In the first term, we observe that the k_5 integral is convergent, and can be evaluated straightaway using Eq. (4.24). The first term then becomes

$$\tfrac{5}{2}\int\frac{k\,d^4k}{(2\pi)^4}2\pi(\sqrt{\phi_c}-1)\ . \tag{4.28}$$

Clearly this is quintically divergent. We define the integral by assuming the domain of integration to be the interior of a large sphere in momentum space: $k\le\Lambda$.

However, general covariance in the four-dimensional space parametrized by x^μ prevents us from assuming that Λ is independent of ϕ_c. We can see this as follows: The coordinate system we are using, in which the background four-dimensional metric is $\phi_c^{-1/3}\delta_{\mu\nu}$, is related to the standard Euclidean-space coordinate system x_0^μ in which the background metric is just $\delta_{\mu\nu}$, by

$$x^\mu=\phi_c^{1/6}x_0^\mu\ . \tag{4.29}$$

Now

$$\int\frac{d^4k}{(2\pi)^4}=\frac{\Lambda^4}{32\pi^4}=\delta^4(0)\ .$$

Thus, under the scaling, Eq. (4.29), we must have

$$\Lambda=\phi_c^{-1/6}\Lambda_0\ , \tag{4.30}$$

where Λ_0 is the corresponding cutoff in the standard coordinate system. Therefore, in evaluating the expression (4.28), we must cut off the first term with $\Lambda=\phi_c^{-1/6}\Lambda_0$, and the second term with Λ_0. This gives

$$\frac{\Lambda_0^5}{8\pi}(\phi_c^{-5/6}\phi_c^{1/2}-1)=\frac{\Lambda_0^5}{8\pi}(\phi_c^{-1/3}-1)\ . \tag{4.31}$$

Finally, returning to Eq. (4.27), we can extract $V_{\text{eff}}(\phi_c)$:

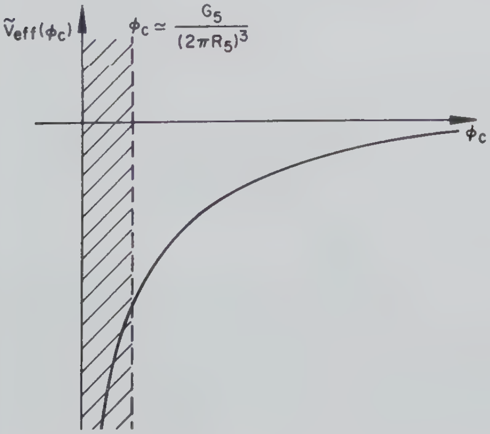


FIG. 1. The one-loop Kaluza-Klein effective potential

$$V_{\text{eff}}(\phi_c)=\frac{\Lambda_0^5}{8\pi}+\frac{5\beta}{(2\pi\phi_c^{1/3}R_5)^5}\ , \tag{4.32}$$

where β is shorthand for the integral appearing in Eq. (4.27). The difference taken in Eq. (4.27) would allow for the presence of an additional constant in

$$\int d^5x\,(g^{(0)})^{1/2}V_{\text{eff}}=\int d^5x\,\phi_c^{-1/3}V_{\text{eff}}$$

but, as noted following Eq. (4.15), this can be absorbed into the normalization of the functional integral. It is physically irrelevant and cleanly separable from the effective potential.

Our last numerical chore is to compute β:

$$\begin{aligned}\beta&=\int\frac{d^4w}{(2\pi)^4}\ln(1-e^{-w})\\&=-\frac{1}{8\pi^2}\sum_{n=1}^{\infty}\frac{1}{n}\int_0^\infty w^3e^{-nw}dw\\&=-\frac{3}{4\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^5}\ .\end{aligned} \tag{4.33}$$

Therefore

$$\begin{aligned}5\beta&=-\frac{15}{4\pi^2}\zeta(5)\\&\cong-0.394.\end{aligned}$$

V. SUMMARY AND DISCUSSION

Our result for the effective potential consists of two pieces: the first is positive, divergent, and independent of the distance L_5 around the fifth dimension; the second is negative, convergent, and depends on L_5 in the manner illustrated in Fig. 1. This state of affairs is reminiscent of the well-known Casimir effect in electrodynamics. There one considers the zero-point energy of the free electromagnetic field between two infinite parallel perfectly conducting plates separated by a distance a, and one finds that the energy per unit volume has the form

$$E(a)=E_0-\frac{c}{a^4}\ . \tag{5.1}$$

Here E_0 is cutoff dependent, but independent of the separation a; the second term is finite and represents a measurable attractive force between the plates. E_0 is just the vacuum energy density of free space, and is to be subtracted out in computing physical quantities.

Our computation is the direct gravitational analog of the usual Casimir energy. The infinite term is an induced cosmological constant, and according to the philosophy put forward in Sec. II, we subtract it from the observable effective potential $\widetilde{V}_{\text{eff}}$:

$$\widetilde{V}_{\text{eff}} = V_{\text{eff}} - \frac{1}{8\pi}\Lambda_0^{\,5} . \tag{5.2}$$

(Of course, if we had chosen to use dimensional continuation or zeta-function regularization, the infinite term would have been defined to zero *ab initio*.) The remaining term is then to be interpreted as an observable energy per unit volume, which tends to make the "distance between the plates"—in this case, the distance around the fifth dimension—contract.

From Fig. 1, however, one sees that there is no apparent limit to the contraction: V_{eff} tends to $-\infty$ as L_5 tends to zero. This is equally true of the Casimir case, but there one expects on physical grounds that a natural cutoff will appear when the structure of the plates becomes important, i.e., when a becomes of order the interatomic separation.

The analogous cutoff in the gravitational case is the Planck length. What we have computed is the first nonvanishing term in a loop expansion, which is an expansion in powers of the reciprocal of the constant multiplying the action, i.e., an expansion in powers of G_5. Now G_5 has dimension $(\text{length})^3$; hence the dimensionless expansion parameter must be

$$\gamma = \frac{G_5}{L_5^{\,3}} \equiv \frac{G_4}{L_5^{\,2}} , \tag{5.3}$$

i.e., the loop expansion should have the form

$$\widetilde{V}_{\text{eff}} = \widetilde{V}^{(1)}_{\text{eff}}(1+\alpha_2\gamma+\alpha_3\gamma^2+\cdots) .$$

Here $\widetilde{V}^{(1)}_{\text{eff}}$ is the one-loop term of Eq. (5.2) and α_n is a dimensionless factor which can be evaluated by computing the finite part of the n-loop contribution to V_{eff}; presumably it is of order 1. Hence the loop expansion will make sense (and our one-loop result will be reliable) only if

$$\gamma \ll 1 , \tag{5.4}$$

i.e.,

$$L_5 \gg (G_4)^{1/2} = L_P ,$$

where L_P is the observed Planck length

$$L_P = 1.6\times 10^{-33}\ \text{cm}. \tag{5.5}$$

Thus our one-loop result loses its validity if $L_5 \lesssim L_P$. We conclude, therefore, that the Casimir force tends to push the fifth dimension down to a size on the order of the Planck length, but we cannot say what happens after that. For the theory to make sense, it must presumably be the case that dynamics at the Planck scale or below will stabilize the extra dimension at this size.

A Planck-size stabilization would have two interesting consequences. First of all, the massive spin-2 modes are charged and their charge is given by a multiple of the ratio of the Planck length to the circumference of the compact dimension. Thus a dynamics which determine this distance relative to the Planck length will, in turn, determine the electric charge of the massive modes. If this charge is to be a multiple of the fine structure constant $\alpha = 1/137$, then the extra dimension would have to be somewhat larger than the Planck length. A stable minimum of the effective potential would also give rise to a mass for the Brans-Dicke scalar, presumably on the order of the Planck mass. This would then screen the scalar out of the effective four-dimensional theory in the same way that finite temperature Debye-screens the long-range electric field in a gauge theory.

The five-dimensional model we have examined in detail is the simplest possible Kaluza-Klein scheme. It invites generalization in a variety of ways. One can, in the first place, have any number d of toroidally compactified dimensions in a D-dimensional space-time. The computation of the effective potential is not significantly more complicated than the $D=5$, $d=1$ case we have considered in this paper. In particular, if one looks at a five-dimensional theory with $d=2$, one can interpret the second compactified dimension as Euclidean time, in which case the distance around it has the interpretation of an inverse temperature β. This case has been analyzed by Rubin and Roth,[24] who find that the attractive Casimir force competes with the thermal pressure: If the ratio β/L_5 is sufficiently large, the fifth dimension will contract; otherwise, it expands. Once can also investigate how the Casimir energy affects $d>1$ compactified spatial dimensions: this will be the subject of a future publication.[25]

Of at least equal importance, and of much greater difficulty, is the problem of extending our analysis to the non-Abelian cases discussed classically in Sec. II. In order to achieve a spontaneously compactified vacuum, it may well be necessary to add additional fields, either ordinary matter[13] or perhaps spinning matter,[14,26] which will give rise to torsion. Even if a sensible classical Kaluza-Klein theory can be constructed, however, the task of performing the mode sum is likely to be very much more complicated.

There is also the problem of incorporating fermions in a realistic way. This may involve the introduction of supersymmetry; the possible close connection between Kaluza-Klein ideas and supersymmetry has been especially emphasized by Witten,[27] and the various possibilities for realizing 11-dimensional supergravity have been discussed by Duff and collaborators.[28]

Finally, we remark that if some version of the Kaluza-Klein idea is right, it must impact on the evolution of the early universe. Some suggestions for cosmology involving Kaluza-Klein theories have already been advanced[9,29]; it remains to be seen whether this approach can be naturally unified with standard big-bang and inflationary cosmological ideas.

APPENDIX A: SOME FEATURES OF FINITE TEMPERATURE GAUGE THEORIES

A system in thermal equilibrium at temperature T is described by the partition function

$$Z=\sum_i e^{-E_i\beta} , \tag{A1}$$

where $\beta=1/T$. Functional integral expressions for the partition function and for thermal expectation values can be derived using standard methods. Euclidean "time" runs between $t=0$ and $t=\beta$, and the fields are either periodic or antiperiodic in this interval depending on whether they describe bosons or fermions. For gauge theories, the Faddeev-Popov gauge-fixing procedure is employed with the ghost field periodic.

The Feynman rules for the quantum loop expansion are formed from the zero-temperature Euclidean Feynman rules by the replacement

$$\int \frac{d^4k}{(2\pi)^4} \to T \sum_{n=\pm\text{integer}} \int \frac{d^3k}{(2\pi)^3} , \tag{A2}$$

where the discrete energy sum is over $k_4=2n\pi T$ for bosons and $k_4=(2n+1)\pi T$ for fermions. The ultraviolet divergences are the same as those of the zero-temperature theory. The zero-temperature counterterms can, therefore, be used to remove the divergences and the temperature T can be used as the scale to define the coupling constant.

At distance scales much larger than $1/T$, the theory simplifies by, in effect, reducing to the same theory in one less dimension. With the external momenta in Green's functions small compared to T, everything except the $n=0$ mode for the bosons can be expected to decouple, leaving only three-dimensional momentum integrals. The only exception to this is the self-mass $\pi_{44}^{ab}(q=0)$ of the A_4 component of the gauge field, which is ultraviolet divergent without the inclusion of the $n\neq 0$ modes. When they are included, one finds

$$\pi_{44}^{ab}(q=0)=(N+N_f/2)\frac{g^2T^2}{3} \tag{A3}$$

for an SU(N) gauge theory with N_f fermions. Thus, a color-electric Debye-screening mass is generated, meaning that at distances much larger than $1/gT$, the fourth (electric) component of the gauge field decouples. With only the three spatial components of the gauge field remaining and with only three-dimensional integrations to be done, the reduction to an effective three-dimensional theory is then complete.

A natural gauge choice for the implementation of this program is the static gauge

$$\partial_4 A_4^a=0 . \tag{A4}$$

It can be obtained by the addition of the gauge-fixing term

$$\Delta\mathscr{L}_{\text{gf}}=\tfrac{1}{2}(\Lambda_\mu A_\mu^a)^2 , \tag{A5}$$

where $\Lambda_\mu=(-i\vec{\Lambda},\lambda\partial_4)$ with $\vec{\Lambda}$ an arbitrary three-vector. The corresponding ghost Lagrangian is

$$\Delta\mathscr{L}_{\text{ghost}}=\eta^{a\dagger}\Lambda_\mu(\partial_\mu\eta^a+gf^{abc}A_\mu^b\eta^c) . \tag{A6}$$

The static gauge is then obtained in the limit $\lambda\to\infty$. It has the following properties:

(1) The A_4 propagator is static ($n=0$ only) with

$$iD_{44}^{ab}(k)=-i\delta^{ab}/k^2 . \tag{A7}$$

At one loop, a mass is generated by the Debye-screening mechanism.

(2) The A_i propagator has a nonstatic contribution ($k_0=2n\pi T,n\neq 0$):

$$n\neq 0\text{: } iD_{ij}^{ab}(k)=-i\delta^{ab}\frac{\delta_{ij}+k_ik_j/k_0}{k^2+{k_0}^2} . \tag{A8}$$

It also has a static ($n=0$) contribution which depends on $\vec{\Lambda}$, that is, on further gauge fixing within the static sector. In the O(3)-covariant gauge $\vec{\Lambda}=-(i/\xi)\vec{\partial}$,

$$n=0\text{: } iD_{ij}^{ab}(\vec{\mathrm{k}})=-i\frac{\delta_{ab}}{\vec{\mathrm{k}}^2}\left[\delta_{ij}+(\xi-1)\frac{k_ik_j}{\vec{\mathrm{k}}^2}\right] . \tag{A9}$$

(3) Only the static ($n=0$) Faddeev-Popov ghost survives. Its propagator in O(3)-covariant gauge is

$$iG^{ab}(\vec{\mathrm{k}})=-i\delta^{ab}\xi/k^2 , \tag{A10}$$

and it couples only to the static A_i propagator [Eq. (A9)], with strength g/ξ.

The static gauge is a natural choice for infrared studies at finite temperature because, from the point of view of the effective three-dimensional theory, it is physical. This can be seen by first noting that the four-dimensional finite-temperature gauge field describes 2 physical degrees of freedom for each value of n. For $n\neq 0$, they correspond to the 2 degrees of freedom of the massive three-vector field $A_i^{(n\neq 0)}$. The propagator for this field appears automatically in the physical, Proca-Wentzel, from [Eq. (A8)]. For $n=0$, the 2 degrees of freedom are accounted for by the massless gauge field $A_i^{(n=0)}$ and the three-dimensional scalar A_4. Each describes 1 physical degree of freedom.

Note added. After this work was completed, we became aware of a paper by An Ing and Chen Shi [proceedings of the Third Marcel Grossman Meeting (unpublished)] in which the same effective potential is computed, for related although somewhat different reasons. We thank E. Witten for bringing this work to our attention.

ACKNOWLEDGMENTS

We thank our colleagues at Yale for many useful comments and for their encouragement. We are especially grateful to Steve Detweiler and Vince Moncrief. We acknowledge the hospitality of the Lewes Center for Physics where some of this work was done. Research was supported in part by the U. S. Department of Energy under Contract No. DE-AC02-76ER03075.

---

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. 966 (1921).

[2] O. Klein, Z. Phys. 37, 895 (1926).

[3] B. DeWitt, in *Lectures at 1963 Les Houches School, Relativity, Groups and Topology,* edited by B. Dewitt and C. DeWitt (Gordon and Breach, New York, 1964), published separately under the title *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); R. Kerner, Ann. Inst. Henri Poincaré 9, 143, (1968), A. Trautman, Rep. Math. Phys. 1, 29 (1970); Y. M. Cho, J. Math. Phys. 16, 2029 (1975); Y. M. Cho and P. G. O. Freund, Phys. Rev. D 12, 1711 (1975); Y. M. Cho and P. S. Jang, *ibid.* 12, 3789 (1975); L. N. Chang, K. I. Macrae, and F. Mansouri, *ibid.* 13, 235 (1976).

[4] C. A. Orzalesi, Fortschr. Phys. 29, 413 (1981); E. Witten, Nucl. Phys. B186, 412 (1981); C. Wetterich, Phys. Lett. 113B, 377 (1982); S. D. Unwin, *ibid.* 103B, 18 (1981); P. G. O. Freund and M. A. Rubin, *ibid.* 97B, 233 (1980).

[5] E. Cremmer and B. Julia, Phys. Lett. 80B, 48 (1978); Nucl. Phys. B159, 141 (1979).

[6] T. Appelquist and A. Chodos, Phys. Rev. Lett. 50, 141 (1983).

[7] H. B. G. Casimir, Proc. K. Ned. Akad. Wet. 51, 793 (1948).

[8] T. Appelquist and R. Pisarski, Phys. Rev. D 23, 2305 (1981); S. Nadkarni, *ibid.* 27, 917 (1983).

[9] A. Chodos and S. Detweiler, Phys. Rev. D 21, 2167 (1980).

[10] J. Scherk and J. H. Schwarz, Nucl. Phys. B153, 61 (1979).

[11] C. Brans and R. H. Dicke, Phys. Rev. 124, 925 (1961); see also S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), Sec. 7.3.

[12] V. Moncrief, private communication.

[13] J. F. Luciani, Nucl. Phys. B135, 111 (1978).

[14] C. A. Orzalesi and M. Pauri, Phys. Lett. 107B, 186 (1981).

[15] C. Wetterich, Ref. 4.

[16] B. S. DeWitt, Phys. Rev. 162, 1195 (1967); 162, 1239 (1967).

[17] G. 't Hooft and M. Veltman, Ann. Inst. Henri Poincaré 20, 69 (1974).

[18] S. Deser and P. van Nieuwenhuizen, Phys. Rev. D 10, 401 (1974); 10, 411 (1974).

[19] A. Salam and J. Strathdee, Ann. Phys. (N.Y.) 141, 316 (1982), Appendix 5.

[20] E. S. Fradkin and G. A. Vilkovisky, Phys. Rev. D 8, 4241 (1973).

[21] B. S. DeWitt, in *General Relativity: An Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, London, 1979), especially Sec. 14.5.2 and pp. 734 and 735.

[22] G. 't Hooft, in *Recent Developments in Gravitation* (Plenum, New York, 1979), especially Sec. 3.

[23] R. Jackiw, Phys. Rev. D 9, 1686 (1974).

[24] M. A. Rubin and B. D. Roth, University of Texas Report No. UTTG-2-83, (unpublished).

[25] T. Appelquist, A. Chodos, and E. Myers, Phys. Lett. (to be published).

[26] M. Duff and C. A. Orzalesi, Phys. Lett. 122B, 37 (1983).

[27] E. Witten, Ref. 4.

[28] M. J. Duff, CERN Report No. TH3451-CERN, 1982 (unpublished) and references therein.

[29] P. G. O. Freund, Nucl. Phys. B209, 146 (1982).

CHARGES FROM EXTRA DIMENSIONS [☆]

Steven WEINBERG
Theory Group, Physics Department, The University of Texas, Austin, TX 78712, USA

Received 31 March 1983

A prescription is given for calculating gauge coupling constants in $(4 + N)$-dimensional models in which N dimensions form an arbitrary compact manifold: Each coupling is given by the ratio of $2\pi(16\pi G)^{1/2}$ to an appropriate root-mean-square circumference of the manifold. It is speculated that these couplings will be of order $1/\sqrt{n}$, where n is the number of matter fields whose one-loop quantum energy density is responsible for the compactification.

It is well known that in the original five-dimensional Kaluza–Klein model [1] [‡1], all electric charges are multiples of the quantity $e = \kappa/R$, where $\kappa^2 \equiv 16\pi G$, and $2\pi R$ is the circumference of the compact fifth dimension. In more general theories [3] based on the compactification of more than one extra dimension, the various Yang–Mills coupling constants are again given by ratios of κ and various circumferences of the compact manifold, but here it is necessary to be precise as to just which coupling constant is related to which circumference. This question acquires physical interest in models where the dynamics of spontaneous compactification impose simple relations among otherwise unrelated circumferences. (For instance, this is the case in supergravity models for which the compact manifold is an Einstein space, composed of a direct product of symmetric spaces [4].) Kaluza–Klein theories already provide an alternative explanation of one of the results of conventional grand unified theories, the quantization of electric charge; perhaps through the dynamical relations among circumferences they can also explain the relations among SU(3) × SU(2) × U(1) couplings as well. This note offers a general prescription [eq. (16)] for calculating gauge coupling constants in terms of appropriate root-mean-square circumferences of the compact subspace.

We consider a $(4 + N)$-dimensional space, with coordinates that can be separated into the coordinates x^μ of ordinary spacetime, plus the coordinates y^n of a compact N-dimensional manifold $\mathcal{M}$. The $(4 + N)$-dimensional metric, including effects of gravity and Yang–Mills fields (but not scalars) in 4 dimensions has the components [‡2]

$$\begin{aligned} \bar{g}_{\mu\nu}(x,y) &= g_{\mu\nu}(x) \\ &\quad + \tilde{g}_{nm}(y)\, \xi^n{}_\alpha(y)\, \xi^m{}_\beta(y)\, A^\alpha{}_\mu(x)\, A^\beta{}_\nu(x)\,, \\ \bar{g}_{\mu n}(x,y) &= \bar{g}_{n\mu}(x,y) = -\tilde{g}_{nm}(y)\, \xi^m{}_\alpha(y)\, A^\alpha{}_\mu(x)\,, \\ \bar{g}_{nm}(x,y) &= \tilde{g}_{nm}(y)\,, \end{aligned} \tag{1}$$

where $g_{\mu\nu}(x)$ is the spacetime metric; $\tilde{g}_{nm}(y)$ is the metric of the compact subspace, assumed to be invariant under a group $\mathcal{G}$ of isometries with Killing vectors $\xi^n{}_\alpha(y)$; and $A^\alpha{}_\mu(x)$ are the gauge fields of this group. At this point, the normalization of the ξ's and A's are not yet separately fixed.

In order to determine the gauge coupling constants, it is necessary to relate the coefficients of the kinematic terms in the Lagrangian for the four-dimensional gauge and gravitational fields. For this purpose, we assume that both of these kinematic terms arise from

---

[☆] Research supported in part by the Robert A. Welch Foundation.

[‡1] For a historical review, see ref. [2].

[‡2] Most of the authors of ref. [3] deal with the special cases where the compact manifold is a group space or a coset space, and give eq. (1) for these cases. Eq. (1) is given in the general case by Witten [3].

the same source – the Einstein lagrangian in $4+N$ dimensions:

$$\bar{\mathcal{L}}_E = -\bar{g}^{1/2}\,\bar{\kappa}^{-2}\,(\bar{R}+\bar{\Lambda})\,, \tag{2}$$

where $\bar{R}$ is the curvature in $4+N$ dimensions, and $\bar{\kappa}$ and $\bar{\Lambda}$ are the $(4+N)$-dimensional gravitational and cosmological constants. A straightforward calculation using (1) in (2) gives the four-dimensional lagrangian [‡3]

$$\begin{aligned}\mathcal{L}_{EYM}(x) &\equiv \int \mathrm{d}^N y\, \bar{\mathcal{L}}_E(x,y)\\ &= -\bar{\kappa}^{-2}\int \mathrm{d}^N y\, \tilde{g}^{1/2}(y)\, g^{1/2}(x)\, [R(x)+\tilde{R}(y)+\bar{\Lambda}\\ &\quad + \tfrac{1}{4}\tilde{g}_{nm}(y)\,\xi^n{}_\alpha(y)\,\xi^m{}_\beta(y)\,F^\alpha{}_{\mu\nu}(x)\\ &\quad \times F^\beta{}_{\lambda\rho}(x) g^{\mu\lambda}(x)\, g^{\nu\rho}(x)]\,. \end{aligned}\tag{3}$$

Here $R(x)$ and $\tilde{R}(y)$ are the scalar curvatures in 4 and N dimensions, respectively, and $F^\alpha{}_{\mu\nu}(x)$ is the usual Yang–Mills curl:

$$F^\alpha{}_{\mu\nu} = \partial_\mu A^\alpha{}_\nu - \partial_\nu A^\alpha{}_\mu - C_{\beta\gamma}{}^\alpha A^\beta{}_\mu A^\gamma{}_\nu\,, \tag{4}$$

where $C_{\beta\gamma}{}^\alpha$ is the real structure constant of the group $\mathcal{G}$ generated by the Killing vectors

$$\xi^m{}_\beta\,\partial_m\,\xi^n{}_\gamma - \xi^m{}_\gamma\,\partial_m\,\xi^n{}_\beta = -C_{\beta\gamma}{}^\alpha\,\xi^n{}_\alpha\,. \tag{5}$$

The four-dimensional gravitational constant $\kappa^2 \equiv 16\pi G$ is now seen to be given by

$$\kappa^{-2} = \bar{\kappa}^{-2}\int \mathrm{d}^N y\, \tilde{g}^{1/2}(y)\,. \tag{6}$$

Inspection of (3) shows that in order for the gauge field $A^\alpha{}_\mu$ to be canonically normalized, the Killing vectors must be selected and normalized so that

$$\langle\xi^n{}_\alpha(y)\,\xi^m{}_\beta(y)\,\tilde{g}_{nm}(y)\rangle = \kappa^2\,\delta_{\alpha\beta}\,, \tag{7}$$

with $\langle\ldots\rangle$ denoting an average over the compact manifold $\mathcal{M}$

$$\langle f(y)\rangle \equiv \int \mathrm{d}^N y\, \tilde{g}^{1/2}(y) f(y) \Big/ \int \mathrm{d}^N y\, \tilde{g}^{1/2}(y)\,. \tag{8}$$

[Also, $\bar{\Lambda}$ must be chosen to cancel $\langle\tilde{R}(y)\rangle$ and any other contributions to the vacuum energy density from other parts of the langrangian.]

Eq. (7) has a useful geometric interpretation. The action of elements g of the group $\mathcal{G}$ on the manifold $\mathcal{M}$ may be parameterized by the matrix $U(g)$ representing g in any *faithful* unitary representation of $\mathcal{G}$:

$$g: y^n \to y^{n'} = Y^n(y, U(g))\,. \tag{9}$$

The Killing vectors are then defined by

$$Y^n(y, 1+\mathrm{i}\epsilon^\alpha t_\alpha) \to y^n + \xi^n{}_\alpha(y)\,\epsilon^\alpha\,, \tag{10}$$

for $\epsilon \to 0$, where t_α are the hermitian generators of $\mathcal{G}$ in the representation $U(g)$, and satisfy a Lie algebra with the same structure constants $C_{\beta\gamma}{}^\alpha$ as in (5) and (4). In particular, an exponential transformation carries a point y_0 along a curve

$$Y^n(y_0, \exp(\mathrm{i}\sigma e^\alpha t_\alpha)) \equiv \mathcal{Y}^n(\sigma; y_0)\,, \tag{11}$$

satisfying the differential equation

$$(\mathrm{d}/\mathrm{d}\sigma)\,\mathcal{Y}^n = \xi^n{}_\alpha(\mathcal{Y})\,e^\alpha\,, \tag{12}$$

with e an arbitrary fixed vector, which for convenience we normalize so that

$$\delta_{\alpha\beta}\,e^\alpha e^\beta = 1\,. \tag{13}$$

If for some e all eigenvalues of $e^\alpha t_\alpha$ are integer multiples of the eigenvalue g_e of lowest non-zero absolute value, then the curve (11) returns to its starting point when σ increases from 0 to $2\pi/g_e$. If furthermore the representation $U(g)$ is N_e-valued for the subgroup generated by $e^\alpha t_\alpha$, then the curve (11) goes N_e times around the manifold $\mathcal{M}$ as σ increases from 0 to $2\pi/g_e$. The circumference of the manifold $\mathcal{M}$ along this curve is therefore

$$\begin{aligned} s(e, y_0) &= N_e^{-1}\\ &\times \int_0^{2\pi/g_e} [\tilde{g}_{nm}(\mathcal{Y})\,\xi^n{}_\alpha(\mathcal{Y})\,\xi^m{}_\beta(\mathcal{Y})\,e^\alpha e^\beta]^{1/2}\,\mathrm{d}\sigma\,. \end{aligned}\tag{14}$$

It is easy to see that the integrand here is actually constant in σ, so the circumference is

$$\begin{aligned} s(e, y_0) &= (2\pi/g_e N_e)\\ &\times [\tilde{g}_{nm}(y_0)\,\xi^n{}_\alpha(y_0)\,\xi^m{}_\beta(y_0)\,e^\alpha e^\beta]^{1/2}\,. \end{aligned}\tag{15}$$

Eq. (7) thus tells us that when we average over the starting points y_0 of the curves (11), the rms circumference $\langle s^2\rangle^{1/2}$ of $\mathcal{M}$ along these curves is $2\pi\kappa/g_e N_e$. That is, the eigenvalue g_e is

---

[‡3] Eq. (3) is given in several of the articles of ref. [3] for the special cases where the compact manifold is a group space or a coset space.

$$g_e = 2\pi\kappa/\langle s^2(e,y)\rangle^{1/2}\, N_e\,, \tag{16}$$

the average taken as in eq. (8).

It remains to pin down the connection between the eigenvalue g_e and the physical gauge coupling constants. For the semi-simple part of the gauge group $\mathcal{G}$, this is already provided for us by eq. (4): the Yang–Mills couplings appear as factors in the structure constants $C_{\beta\gamma}{}^{\alpha}$ of the algebra t_α, and are therefore proportional to the g_e. More specifically, if the adjoint representation is faithful [‡4], then we can use $-\mathrm{i}C_{\alpha\beta}{}^{\gamma}$ as our generator $(t_\alpha)_\beta{}^\gamma$, so the lowest non-zero eigenvalue of the matrix $-\mathrm{i}e^\alpha G_{\alpha\beta}{}^\gamma$ is just g_e, given by eq. (16). However, this reasoning clearly tells us nothing about the physical significance of g_e when $e^\alpha t_\alpha$ is the generator of a U(1) subgroup of $\mathcal{G}$.

In order to understand the physical significance of all g_e for non-semi-simple as well as semi-simple gauge groups, it is useful to introduce additional matter fields. Consider a complex scalar field $\phi(x,y)$ in 4 + N dimensions. It can be expanded in a set of functions $u^l{}_\nu(y)$ that furnish unitary irreducible representations U^l of $\mathcal{G}$:

$$\phi(x,y) = \sum_{l,r} u_r{}^l(y)\,\phi_r{}^l(x)\,, \tag{17}$$

$$\xi^m{}_\alpha(y)\,\partial_m u_r{}^l(y) = \mathrm{i}\sum_s (t_\alpha{}^l)_{rs} u_s{}^l(y)\,, \tag{18}$$

$$\int u_r{}^k(y)\,{}^*u_s{}^l(y)\,\tilde{g}^{1/2}(y)\,\mathrm{d}y = \delta_{kl}\delta_{rs}\,. \tag{19}$$

By applying (18) to (5) and (19), it is easy to see that the $t_\alpha{}^l$ are hermitian and satisfy the Lie algebra with structure constants $C_{\alpha\beta}{}^\gamma$. The kinematic lagrangian for the scalar in 4 + N dimensions is taken as

$$\bar{\mathcal{L}}_\phi = -\bar{g}^{1/2}\bar{g}^{NM}\,(\partial_N\phi)^*\,(\partial_M\phi)\,, \tag{20}$$

and in four dimensions, this gives

$$\mathcal{L}_\phi \equiv \int \mathrm{d}^N y\,\bar{\mathcal{L}}_\phi = -g^{1/2}\,g^{\mu\nu}\sum_{r,l}(\mathrm{D}_\mu\phi^l)_r^*(\mathrm{D}_\nu\phi^l)_r\,, \tag{21}$$

with

$$(\mathrm{D}_\mu\phi^l)_r = \partial_\mu\phi_r{}^l + \mathrm{i}\sum (t_\alpha{}^l)_{rs}\,\phi_s{}^l A^\alpha{}_\mu\,. \tag{22}$$

Along the curve (11), eq. (18) gives $u^l(\mathcal{Y})$ proportional to $\exp(\mathrm{i}\sigma e^\alpha t_\alpha{}^l)$, so since $\mathcal{Y}$ returns to its starting point when σ increases from 0 to $2\pi/g_e$, the eigenvalues of $e^\alpha t_\alpha{}^l$ in a (not necessarily faithful) $N_e{}^l$-valued representation U^l of $\mathcal{G}$ must be integer multiples of $g_e/N_e{}^l$. Eq. (22) makes it clear that these eigenvalues are the physical coupling constants that determine the interaction of these matter fields with the gauge fields $A^\alpha{}_\mu$.

To see how this works in a specific example, consider the homogeneous isotropic space S_N in N dimensions, for which the gauge group $\mathcal{G}$ is SO(N + 1). We parameterize points in this space by N + 1 euclidean coordinates y^n, subject to the constraint

$$\sum_n (y^n)^2 = R^2\,.$$

For the Killing curve through y generated by rotations in the n-m plane, the metric circumference is obviously

$$s(e,y) = 2\pi[(y^n)^2 + (y^m)^2]^{1/2}\,,$$

The average values of all $(y^k)^2$ are equal, and their sum is R^2, so

$$\langle (y^n)^2\rangle = \langle (y^m)^2\rangle = R^2/N+1\,,$$

and therefore the mean square circumference is

$$\langle s^2\rangle = 8\pi^2\,R^2/N+1\,.$$

If we define the coupling constant g so that in the single-valued faithful defining representation of SO (N + 1) the eigenvalues of $e^\alpha t_\alpha$ are $g, -g, 0, 0, \ldots$, then (16) gives

$$g = (\kappa/Ry)[\tfrac{1}{2}(N+1)]^{1/2}\,.$$

This is independent of e^α because the space S_N is isotropic.

For one example of the determination of gauge coupling constants for an anisotropic compact manifold, consider the four-dimensional manifold CP_2, whose isometry group is SU(3). The coordinates can be taken as the pair of complex numbers z_1/z_3 and z_1/z_3, where z_1, z_2, and z_3 are a complex triplet transforming according to the defining representation of SU(3). These ratios can be expressed in terms of 4 real angular variables [5]

[‡4] Note that the adjoint representation can be non-faithful (faithless?) as for example where the compact manifold is the group space of SU(2), with isometry group SU(2) [not SO(3)]; in the adjoint representation of SU(2) the 2 × 2 matrix −1 is represented by the 3 × 3 matrix +1. In this case the lowest eigenvalue of $-\mathrm{i}e^\alpha C_{\alpha\beta}{}^\gamma$ has twice the value given by (16).

$$z_1/z_3 = \tan\chi \cos\tfrac{1}{2}\theta \exp[\mathrm{i}\tfrac{1}{2}(\phi + \psi)] ,$$

$$z_2/z_3 = \tan\chi \sin\tfrac{1}{2}\theta \exp[\mathrm{i}\tfrac{1}{2}(\phi - \psi)] ,$$

with ranges

$$0 \leqslant \chi \leqslant \tfrac{1}{2}\pi, \quad 0 \leqslant \theta \leqslant \pi,$$

$$0 \leqslant \phi \leqslant 2\pi, \quad -2\pi \leqslant \psi \leqslant 2\pi .$$

The non-vanishing elements of the metric are

$$\tilde{g}_{\chi\chi} = R^2 ,$$

$$\tilde{g}_{\theta\theta} = \tfrac{1}{4}R^2 \sin^2\chi ,$$

$$\tilde{g}_{\psi\psi} = \tfrac{1}{4}R^2(1\text{-}\sin^2\chi \cos^2\theta) \sin^2\chi ,$$

$$\tilde{g}_{\psi\phi} = \tfrac{1}{4}R^2 \cos^2\chi \sin^2\chi \cos\theta ,$$

$$\tilde{g}_{\phi\phi} = \tfrac{1}{4}R^2 \cos^2\chi \sin^2\chi ,$$

where R is an arbitrary scale parameter. The Killing curves produced by the standard SU(3) generators t_3 and t_8 are just the trajectories with ψ or ϕ varying (respectively) with other angular variables held fixed, so the circumferences here are

$$s_3(\chi, \theta, \phi) = 4\pi(\tilde{g}_{\psi\psi})^{1/2}$$

$$= 2\pi R \sin\chi (1 - \sin^2\chi \cos^2\theta)^{1/2} ,$$

$$s_8(\chi, \theta, \psi) = 2\pi(\tilde{g}_{\phi\phi})^{1/2}$$

$$= \pi R \cos\chi \sin\chi .$$

The rms circumferences, evaluated with weight $\tilde{g}^{1/2} \propto \sin^3\chi \cos\chi \sin\theta$, are then

$$\langle s_3^2\rangle^{1/2} = \sqrt{2}\pi R, \quad \langle s_8^2\rangle^{1/2} = 6^{-1/2}\pi R .$$

As ψ runs from -2π to $+2\pi$, the z_i return to their original values, so the triplet representation is single-valued for the subgroup generated by t_3. Hence if we define coupling constant g so that the eigenvalues of t_3 in the triplet representation are g, $-g$, and 0, then (16) gives

$$g = \sqrt{2}\,\kappa/R .$$

On the other hand, as ϕ runs from 0 to 2π, the phases of z_1, z_2, and z_3 change by $\pi/3$, $\pi/3$, and $-2\pi/3$, so the triplet representation is six-valued for the subgroup generated by t_8. Hence if we define a coupling constant g' so that the eigenvalues of t_8 in the triplet representation are

$$g' = (2/3)^{1/2}\,\kappa/R .$$

The relation $g'/g = 1/\sqrt{3}$ is of course just what is required by the simple group SU(3).

The spaces S_N and CP_2 are all Einstein spaces, for which

$$\tilde{R}_{nm} = -a^{-2}\,\tilde{g}_{nm} .$$

The parameters R and a are related by

$$R/a = (N-1)^{1/2}, \quad \text{for } S_N ,$$

$$R/a = \sqrt{6}, \quad \text{for } CP_2 .$$

Hence if the whole compact manifold is an Einstein space, and is composed of a direct product of various S_N's and CP_2's, then the coupling constants of the associated gauge groups (normalized as above) are

$$g = (\kappa/a)\,[(N+1)/2(N-1)]^{1/2}, \quad \text{for } SO(N+1) ,$$

$$g = (\kappa/a)/\sqrt{3}, \quad \text{for } SU(3) .$$

The ratios of squared coupling constants are rational numbers, but otherwise there does not seem to be any relation between these ratios and those that would be dictated by embedding these gauge groups in a simple group. On the other hand, it is possible that the gauge invariance associated with some simple isometry group of the compact manifold could be spontaneously broken by vacuum expectation values, that for some reason do not affect the compact manifold itself. In this case coupling ratios (such as the ratio $g'/g = 1/\sqrt{3}$ derived above for CP_2) may persist despite the breakdown of the simple gauge group.

We close with a speculation. The known gauge coupling constants, extrapolated to an energy of order $1/\kappa$, are rather less than unity. In the context of Kaluza–Klein theories, this implies that the size R of the compact manifold is somewhat larger than the Planck length κ. Indeed it is necessary to assume that R is larger than κ in order to justify perturbative calculations of spontaneous compactification. But why should R and κ not be of the same order of magnitude? The answer may have to do with the number of matter fields responsible for the compactification. For a manifold of typical circumference R, we would expect the oneloop vacuum free energy density [‡5] of n types of massless matter field in $4 + N$ dimensions to be very roughly of order nR^{-4-N}. The Einstein tensor is of order $1/R^2$ and the $(4 + N)$-dimensional gravitational constant $\bar{\kappa}^2$ is of order $\kappa^2 R^N$, so Einstein's equation gives $1/R^2$ of order $n\kappa^2/R^4$; that is, κ^2/R^2 is of order $1/n$, and typical gauge coupling constants are of order $1/\sqrt{n}$. Both the one-loop graviton contribution to the vacuum energy density and the two-loop gravitational and gauge-boson corrections to the matter energy density then produce only small shifts in κ/R and the gauge coupling constants, suppressed by an additional factor of 1/n.

I am grateful for stimulating conversations with Michael Duff, Christopher Pope and Mark Rubin.

[‡5] The one-loop contribution to the vacuum free energy density has been calculated for the original Kaluza–Klein theory in ref. [6]. In this case the curvature of the compact manifold vanishes, and so equilibrium is impossible unless we add massive matter fields (of spin 1/2 as well as spin 0), to make the vacuum free energy density itself stationary, as in ref. [7]. Theories of the sort proposed here are only possible in 6 or more dimensions.

References

[1] Th. Kaluza, Sitz. Preuss. Akad. Wiss. Kl. (1921) 966;
O. Klein, Z. Phys. 37 (1926) 895;
H. Mandel, Z. Phys. 39 (1926) 136;
O. Klein, Nature 118 (1926) 516;
A. Einstein and P. Bergmann, Ann. Math. 39 (1938) 683.

[2] A. Pais, Subtle is the Lord ... (Oxford U.P., New York, 1982) pp. 329–336.

[3] B. DeWitt, in: Relativity, groups, and topology (Gordon and Breach, New York, 1964) p. 725;
J. Scherk and J.H. Schwarz, Phys. Lett. 57B (1975) 463;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
E. Cremmer and J. Scherk, Nucl. Phys. B108 (1976) 408;
P.G.O. Freund and M. Rubin, Phys. Lett. 97B (1980) 233;
E. Witten, Nucl. Phys. B186 (1981) 412;
E. Cremmer, in: Supergravity 81, eds. S. Ferrara and J. Taylor (Cambridge U.P., London, 1982);
M. Duff, in: Supergravity 81, eds. S. Ferrara and J. Taylor (Cambridge U.P., London, 1982);
A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316;
M.J. Duff and D. Toms, in: Unification of fundamental interactions II, eds. S. Ferrara and J. Ellis (Plenum, New York, 1982);
M.J. Duff and C.N. Pope, Imperial College preprint ICTP/82-83/7 (1983).

[4] M.J. Duff and D. Toms, in: Unification of fundamental interactions II, eds. S. Ferrara and J. Ellis (Plenum, New York, 1982).

[5] G.W. Gibbons and C.N. Pope, Commun. Math. Phys. 61 (1978) 239.

[6] T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141.

[7] M. Rubin and B.D. Roth, Texas preprint UTTG-2-83 (1983).

TEMPERATURE EFFECTS IN FIVE-DIMENSIONAL KALUZA–KLEIN THEORY*

Mark A. RUBIN and Bernard D. BARKAN ROTH

Theory Group, Department of Physics, The Univerisity of Texas at Austin, Austin, Texas 78712, USA

Received 26 January 1983
(Corrected version received 25 May 1983)

We examine the combined effects of thermal and quantum-mechanical fluctuations in a Kaluza–Klein universe with a single compact spatial dimension of length $\bar{L}_5$. From the one-loop effective potential for $\bar{L}_5$ two types of instability are evident. If initially $\bar{L}_5$ is less than a temperature-dependent critical length, $\bar{L}_5$ will shrink at least down to a length comparable to the Planck length. This instability has been noted by Appelquist and Chodos in the zero-temperature limit. If, on the other hand, $\bar{L}_5$ starts out larger than the critical length, it will tend to increase. This result suggests that temperature effects may play an important role in Kaluza–Klein cosmological models.

Kaluza–Klein theory [1] is most appealing when the "extra" dimensions are regarded, not simply as a mathematical stratagem, but rather as a physical reality, with small but finite lengths. Can we ever hope to find direct evidence that this viewpoint is correct? Since at present the lengths of the extra dimensions are at most one or two orders of magnitude larger than the Planck length [2] it is difficult to point to present-day phenomena which might signal their existence. If in fact the non-zero size of the extra dimensions has any role in physics, it seems likely that this role was played in the high-temperature environment of the very early universe. With this motivation we shall examine the role of temperature in a toy Kaluza–Klein model.

The original and simplest version of a Kaluza–Klein universe is the five-dimensional spacetime consisting of the direct product of four-dimensional Minkowski spacetime M^4 with the circle S^1. The classical ground state of this theory corresponds to the metric

$$g_{\mu\nu} = \mathrm{diag}\,(-1, 1, 1, 1, g_{55})\,, \tag{1}$$

where g_{55} is a constant. If the coordinate x^5 which measures distance around the compact dimension runs from 0 to $2\pi R_5$, the physical circumference of the fifth dimension is seen to be

$$L_5 \equiv 2\pi (g_{55})^{1/2} R_5\,. \tag{2}$$

---

* Supported in part by the Robert A. Welch Foundation.

For any value of L_5, this spacetime is a stable solution of the five-dimensional vacuum Einstein equation. This is the classical state of affairs; what happens when quantum effects are included?

At the quantum level, the requirement that the metric $g_{\mu\nu}$ satisfy the classical Einstein equation is replaced by the requirement that $\bar{g}_{\mu\nu}$, the expectation value of $g_{\mu\nu}$, minimize the effective action functional $\Gamma[\bar{g}_{\mu\nu}]$. If $\bar{g}_{\mu\nu}$ is of the form (1), the effective action is an ordinary function of the parameter $\bar{g}_{55}$, or equivalently, of $\bar{L}_5 = 2\pi(\bar{g}_{55})^{1/2}R_5$. The effective potential in the $O(\hbar)$ (one-loop) approximation has been computed by Appelquist and Chodos [3], who show that it drops monotonically to $-\infty$ at $\bar{L}_5 = 0$.

This seems to indicate that the compact dimension can have no size other than zero. In fact, we can only conclude that the compact dimension will tend to shrink until its circumference is comparable to the Planck length, L_P; below this length the one-loop approximation loses its validity, and we can say nothing with certainty.

In this paper we wish to point out that, if the calculation described above is modified so as to be valid for non-zero temperature, the effective potential drops to $-\infty$ not only at $\bar{L}_5 = 0$, *but also at* $\bar{L}_5 = \infty$. The turning point between positive and negative slope occurs at $\bar{L}_5$ approximately 1.4 times the inverse temperature β. This conclusion holds for values of $\bar{L}_5$ and β such that $\bar{L}_5 \gg L_P$ and $\beta \gg L_P$, since it is only for this range of parameters that the one-loop result is a good approximation*. The effective potential at constant temperature β^{-1}, $\tilde{\Gamma}$, is plotted as a function of $\bar{L}_5/\beta$ in fig. 1.

The above result implies that, under isothermal conditions, the model Kaluza–Klein universe we are considering here is unstable in two modes. At temperature β^{-1}, the circumference of the compact dimension will shrink at least down to $O(L_P)$

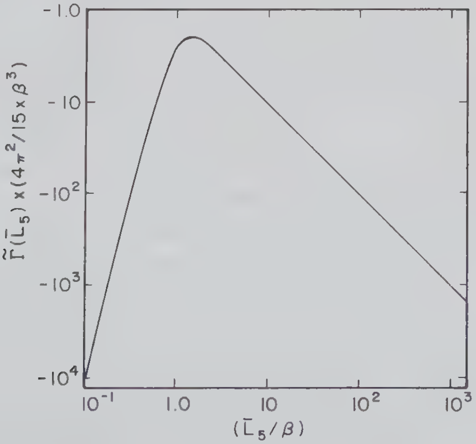


Fig. 1. The effective potential at fixed temperature as a function of the circumference of the fifth dimension.

---

* In addition, if $\bar{L}_5/\beta \gtrsim (\beta/L_P)^{2/3}$, the calculation presented here must be modified to take into account curvature induced by thermal stress-energy. See appendix.

if it is initially smaller than approximately $1.4\,\beta$; it will expand if it has started out larger than $\sim 1.4\beta$. (There is, of course, a point of unstable equilibrium at exactly the critical circumference.) The same qualitative features are found at fixed entropy. The details of these calculations are described in the appendix.

The behavior of this system can be understood physically through the Casimir effect analogy used by Appelquist and Chodos. At zero temperature, the gravitational zero-point energy of the Kaluza–Klein ground state leads to the collapse of the fifth dimension in direct correspondence to the way in which two parallel conducting plates in M^4 are drawn together by the negative zero-point energy of the electromagnetic field. (Boundary conditions in the Casimir problem yield the topology S^1 for the coordinate normal to the plate.) If a gas of photons at fixed temperature is introduced between the plates, the net pressure on the plates will be the sum of two contributions: the positive pressure from thermal photons, of constant magnitude; and the negative Casimir pressure, varying in inverse proportion to the fourth power of the plate separation. If the plates start out close together, the Casimir pressure is stronger than that of the thermal photons and the plates collapse. If they start out at such a distance that the Casimir pressure is weaker, they will recede to infinite separation. Qualitatively identical behavior will be observed under adiabatic conditions. The thermal photon pressure will change as the plate separation changes, but only as (separation) $^{4/3}$; i.e. much more slowly than the Casimir pressure.

It is by no means straightforward to extrapolate from the above calculation to a more realistic Kaluza–Klein model. Such a model would remove the artificial constraints of a non-dynamical external geometry and a time-independent internal geometry. It would also involve more than just one compact dimension (seven is a currently popular number [4]), with compactification brought about by vacuum expectation values for non-gravitational fields, and would include fermionic fields. The presence of curvature in both the compact and non-compact dimensions, the response of the VEV's to changes in temperature, and fermion degeneracy pressure might all contribute to behavior very different from that observed in the present case. We are currently studying models incorporating the features mentioned above, and hope to report on this work in the near future [9].

Our results suggest a mechanism which, in a realistic Kaluza–Klein cosmological model, could cause dimensions which are initially comparable in scale to evolve in radically different ways, some remaining small (or actually decreasing in size), while others expand, perhaps indefinitely. They also suggest the possibility of a sort of "phase transition", in which compact dimensions initially expand to large scales, only to contract suddenly when the universe cools below a certain temperature. If the internal dimensions were at one time much larger than at present, they may well have left observational relics.

We would like to thank Steven Weinberg for suggesting to us that Kaluza–Klein theory at non-zero temperature would be a fruitful area of investigation. To him, and to Henry Kandrup, Bengt Nilsson, Andrew Redfield, and Xerxes Tata, thanks are also due for very helfpful discussions. We also wish to thank David Deutsch for programming the HP-41C. BDBR would like to thank the participants of the FLAG Seminar.

Appendix

COMPUTATION OF THE EFFECTIVE POTENTIAL

In computing the $O(\hbar)$ part of the effective action, we shall only consider those contributions due to quantum and thermal fluctuations which are near the classical solution (1). At least one other stationary point of the euclidean action may be of significance for the system we are studying; namely, the Witten instanton [5], which leads to the effective destruction of the spacetime. Nevertheless, we feel that it is important to study, in and of themselves, the effects caused by fluctuations about $M^4 \otimes S^1$, ignoring for the present the possibile instanton-induced instability. This is because (as pointed out in footnote 15 of [3]) fluctuations about the classical groundstate will play a role in *any* Kaluza–Klein model. What is learned about the effect of these fluctuations in the five-dimensional case is therefore of potential relevance for realistic models. On the other hand, it is by no means clear whether the existence of semiclassical instabilities of the Witten type is a general feature of Kaluza–Klein theories, or, rather, is only a property of certain models. (Even in the $M^4 \otimes S^1$ case the alleged instability is only in fact present if one grants the assumption that quantum gravity permits changes in topology [5].) Models which do possess such instabilities, and which cannot be stabilized in some manner, must of course be excluded as candidates for realistic Kaluza–Klein theories. Suitable generalizations of the analysis presented here would then apply to those models which do not suffer from instabilities of the Witten type.

The one-loop approximation to the effective action for $\bar{g}_{\mu\nu}$ at temperature β^{-1} is then given by*

$$\Gamma[\bar{g}_{\mu\nu}] = S[\bar{g}_{\mu\nu}] + \tfrac{1}{2}\ln\det S_2[\bar{g}_{\mu\nu}] - \ln\det \Delta[\bar{g}_{\mu\nu}]\,, \tag{3}$$

where

$$S = S_E + S_{GF}\,. \tag{4}$$

S_E is the five-dimensional euclidean Einstein action

$$S_E = \frac{1}{16\pi G_5}\int d^5X\,\sqrt{g}\,R\,. \tag{5}$$

$\bar{g}$ is the determinant of $\bar{g}_{\mu\nu}$; R is the Ricci curvature scalar constructed from $\bar{g}_{\mu\nu}$, and G_5 is the five-dimensional Newton constant with dimension of $(\text{length})^3$. S_{GF}

---

* For the effective action formalism at zero temperature see, for example, ref. [6a]; for finite temperature field theory see ref. [6b].

is the gauge-fixing action, for which we choose

$$S_{\mathrm{GF}}=-\frac{1}{32\pi G_5}\int \mathrm{d}^5x\,\sqrt{g}\,g_{\sigma\tau}g^{\sigma\mu}{}_{,\mu}g^{\tau\nu}{}_{,\nu} \tag{6}$$

corresponding to the gauge condition $\partial_\mu g^{\mu\nu}=0$. The DeWitt–Faddeev–Popov matrix Δ corresponding to (6) is

$$\Delta(x,y)_\beta{}^\nu=-(\partial^\nu\,\partial_\beta+g^\nu{}_\beta\partial^\sigma\,\partial_\sigma)\delta^5(x-y)\,. \tag{7}$$

We define the operator

$$S_2^{\mu\nu\alpha\beta}(x,y)\equiv\left.\frac{\delta^2S}{\delta g_{\mu\nu}(x)\,\delta g_{\alpha\beta}(y)}\right|_{g_{\sigma\tau}=\bar{g}_{\sigma\tau}}\,. \tag{8}$$

$S_2^{\mu\nu\alpha\beta}(x,y)$ is to be viewed as an infinite-dimensional matrix S_{2AB} where different values of the "index" $A(B)$ correspond to different values of μ, ν, and x (α, β and y). For example, the eigenvalue equation is

$$\int \mathrm{d}^5y\,S_{2\mu\nu}{}^{\alpha\beta}(x,y)v_{\alpha\beta}(y)=\Lambda v_{\mu\nu}(x)\,, \tag{9}$$

where $v_{\mu\nu}(x)$ is the "eigenvector"corresponding to Λ.

We evaluate $\Gamma[\bar{g}_{\mu\nu}]$ for euclidean $\bar{g}_{\mu\nu}$ of the form

$$\bar{g}_{\mu\nu}=\mathrm{diag}\,(1,1,1,1,\bar{g}_{55})\,. \tag{10}$$

As in eq. (1) $\bar{g}_{55}$ is here a constant, and we may define the finite-temperature effective potential

$$\tilde{\Gamma}(\bar{L}_5)\equiv\Gamma[\bar{g}_{\mu\nu}]\Big/\int \mathrm{d}^3\boldsymbol{x}\,, \tag{11}$$

where

$$\bar{L}_5\equiv 2\pi(\bar{g}_{55})^{1/2}R_5\,. \tag{12}$$

The determinants which appear in the effective action (3) arise from path integration over fluctuations about the classical background (10) at temperature β^{-1}. Because we are considering a field theory at non-zero temperature, the fluctuations must be periodic in euclidean time with period β. Because they are defined on a manifold with topology $\mathbb{R}^4\otimes\mathrm{S}^1$, they must be periodic in the x^5 coordinate with period $2\pi R_5$. Therefore the determinants are to be evaluated in the space of tensor fields on $\mathbb{R}^3\otimes\mathrm{S}^1\otimes\mathrm{S}^1$ which are periodic in x^4 with period β, and periodic in x^5 with period $2\pi R_5$.

To compute the regularized determinant of an operator, say $M(x-y)$, we make use of the ζ-function technique [7]:

$$\ln\det M(x-y)=-\zeta'_M(0)\,. \tag{13}$$

The generalized zeta-function $\zeta_M(s)$ is defined by the relation

$$\zeta_M(s) = \sum_n \Lambda_n^{-s}, \tag{14}$$

for s sufficiently large that the sum converges; Λ_n is the nth eigenvalue of M, the sum being taken over all non-zero Λ_n. For other values of s, $\zeta_M(s)$ is defined by the analytic continuation of (14). From eq. (14) if follows that

$$\zeta_M(s) = \sum_l \sum_{m,n=-\infty}^{\infty} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} [\Lambda_l(\boldsymbol{k}, m, n)]^{-s} \cdot \int \mathrm{d}^3\boldsymbol{x}, \tag{15}$$

where $\Lambda_l(\boldsymbol{k}, m, n)$ is the lth eigenvalue of the Fourier-transformed matrix $M(\boldsymbol{k}, m, n)$,

$$M(\boldsymbol{k}, m, n) = \int_{-\infty}^{\infty} \mathrm{d}^3\boldsymbol{x} \int_0^{\beta} \mathrm{d}x^4 \int_0^{2\pi R_5} \mathrm{d}x^5 \exp\left(-\mathrm{i}\left(\boldsymbol{k}\cdot\boldsymbol{x} + \frac{2\pi m}{\beta}x^4 + \frac{n}{R_5}x^5\right)\right) M(x). \tag{16}$$

(Boldface indicates three-vectors in the non-compact space.)

For metrics of the form (10),

$$S_{2\mu\nu}{}^{\alpha\beta}(k) = \frac{1}{32\pi G_5}\sqrt{\bar{g}_{55}}[(\bar{g}_{\mu\nu}\bar{g}^{\alpha\beta} - \bar{g}_{\mu}^{(\alpha}\bar{g}_{\nu}^{\beta)})k^{\sigma}k_{\sigma} - \bar{g}^{\alpha\beta}k_{\mu}k_{\nu} - \bar{g}_{\mu\nu}k^{\alpha}k^{\beta}], \tag{17}$$

$$\Delta_{\beta}{}^{\nu}(k) = k_{\beta}k^{\nu} + \bar{g}_{\beta}{}^{\nu}k^{\sigma}k_{\sigma}, \tag{18}$$

where

$$k_{\mu} \equiv \left(\boldsymbol{k}, \frac{2\pi m}{\beta}, \frac{n}{R_5}\right). \tag{19}$$

The eigenvalues of the matrices (17) and (18) are all of the form

$$\Lambda_l(k) = k^{\sigma}k_{\sigma}\lambda_l \tag{20}$$

where λ_l is a number independent of k_{μ}. Using (16), we see that the zeta functions for S_2 and Δ are, respectively,

$$\zeta_{S_2}(s) = \hat{\zeta}(s) \sum_{l=1}^{15} \lambda_{S_2 l}^{-s} \int \mathrm{d}^3\boldsymbol{x}, \tag{21}$$

$$\zeta_{\Delta}(s) = \hat{\zeta}(s) \sum_{l=1}^{5} \lambda_{\Delta_l}^{-s} \int \mathrm{d}^3\boldsymbol{x}, \tag{22}$$

where

$$\begin{aligned}\hat{\zeta}(s) &= \sum_{m,n=-\infty}^{\infty} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} (k^{\sigma}k_{\sigma})^{-s} \\ &= \sum_{m,n=-\infty}^{\infty} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \left(\boldsymbol{k}\cdot\boldsymbol{k} + \frac{4\pi^2 m^2}{\beta^2} + \frac{4\pi^2 n^2}{\bar{L}_5^2}\right)^{-s}\end{aligned} \tag{23}$$

$$= \frac{1}{2\pi^2}\left[\int_0^\infty \mathrm{d}l\, l^{2-2s} + \sum_{m,n=-\infty}^{\infty}{}' \int_0^\infty \mathrm{d}l\, l^2\left(l^2 + \frac{4\pi^2 m^2}{\beta^2} + \frac{4\pi^2 n^2}{\bar{L}_5^2}\right)^{-s}\right], \tag{24}$$

where the prime on the sum indicates that the $n = m = 0$ term is excluded. The first term in (24) is infrared-divergent for large s, but this is a spurious divergence and can be discarded. (See p. 138 of [7].) Integrating the remaining term,

$$\hat{\zeta}(s) = \frac{1}{8\pi^{3/2}} \frac{\Gamma(s-\frac{3}{2})}{\Gamma(s)} \sum_{m,n=-\infty}^{\infty}{}' \left[(am)^2 + (bn)^2\right]^{-(s-3/2)}, \tag{25}$$

where

$$a \equiv \frac{2\pi}{\beta}, \qquad b \equiv \frac{2\pi}{\bar{L}_5}. \tag{26}$$

The sum in (25) is a type of Epstein zeta function of order 2, Z_2, and converges absolutely for $s > \frac{5}{2}$ [8]:

$$Z_2(a, b; r) \equiv \sum_{m,n=-\infty}^{\infty}{}' \left[(am)^2 + (bn)^2\right]^{-r}, \qquad r > 1. \tag{27}$$

So,

$$\hat{\zeta}(s) = \frac{1}{8\pi^{3/2}} \frac{\Gamma(s-\frac{3}{2})}{\Gamma(s)} Z_2(a, b; s-\tfrac{3}{2}). \tag{28}$$

For $r < 1$, $Z_2(a, b; r)$ is defined by analytic continuation. We may obtain a series expression for $Z_2(a, b; r)$ for the case $r < 1$ using the reflection formula [8], which yields, for $r = s - 3/2$,

$$\pi^{-(s-\frac{3}{2})}\Gamma(s-\tfrac{3}{2})Z_2(a, b; s-\tfrac{3}{2}) = (ab)^{-1}\pi^{-(\frac{5}{2}-s)}\Gamma(\tfrac{5}{2}-s)Z_2(a^{-1}, b^{-1}; \tfrac{5}{2}-s). \tag{29}$$

So,

$$\hat{\zeta}(s) = \frac{1}{8\pi^{\frac{11}{2}-2s}} \frac{1}{ab} \frac{\Gamma(\frac{5}{2}-s)}{\Gamma(s)} Z_2(a^{-1}, b^{-1}; \tfrac{5}{2}-s). \tag{30}$$

We see that

$$\hat{\zeta}(0) = 0. \tag{31}$$

since $\Gamma(s)$ diverges like $1/s$ as $s \to 0$, while the other factors in (30) are regular at $s = 0$.

Finally,

$$\hat{\zeta}'(0) = \frac{1}{8\pi^{11/2}} \frac{\Gamma(\frac{5}{2})}{ab} Z_2(a^{-1}, b^{-1}; \tfrac{5}{2}). \tag{32}$$

Eqs. (21), (22), and (31) show that

$$\zeta'_{s_2}(0) = 15\hat{\zeta}'(0)\int \mathrm{d}^3\boldsymbol{x}, \tag{33}$$

$$\zeta'_{\Delta}(0)=5\hat{\zeta}'(0)\int \mathrm{d}^3\boldsymbol{x}\,. \tag{34}$$

Since $S[\bar{g}_{\mu\nu}]=0$ for $\bar{g}_{\mu\nu}$ of the form (10), we find from eqs. (3), (11)–(13), and (32)–(34) that

$$\tilde{\Gamma}(L_5)=-\frac{15}{4\pi^2}\frac{1}{\beta^3}f\left(\frac{\bar{L}_5}{\beta}\right)=-\frac{15}{4\pi^2}\frac{1}{\bar{L}_5^3}f\left(\frac{\beta}{\bar{L}_5}\right), \tag{35}$$

where f is the function

$$f(y)\equiv y^{-4}\zeta(5)+y[\zeta(5)+2S_{5/2}(y)]\,, \tag{36}$$

$$S_{5/2}(y)\equiv\sum_{m,n=1}^{\infty}(m^2+y^2n^2)^{-5/2}\,; \tag{37}$$

$\zeta(5)$ above is the usual Riemann zeta function. For $\bar{L}_5\ll\beta$,

$$\tilde{\Gamma}(\bar{L}_5)\sim-\frac{15\zeta(5)}{4\pi^2}\frac{\beta}{\bar{L}_5^4}, \tag{38}$$

while for $\bar{L}_5\gg\beta$,

$$\tilde{\Gamma}(\bar{L}_5)\sim-\frac{15\zeta(5)}{4\pi^2}\frac{\bar{L}_5}{\beta^4}, \tag{39}$$

$\tilde{\Gamma}(\bar{L}_5)$ is plotted in fig. 1.

(Note that it would be *incorrect* in eq. (11) to divide out a factor of the total spacetime volume in the definition of the effective potential, i.e. to define $\tilde{\Gamma}(\bar{L}_5)=\Gamma(\bar{g}_{\mu\nu})/\int \mathrm{d}^5x\,\sqrt{\bar{g}}$. The quantity which must be minimized as a function of $\bar{L}_5$ is $\Gamma(\bar{L}_5)$. The three-volume $\int \mathrm{d}^3\boldsymbol{x}$ is simply an infinite constant, so the minimum of $\tilde{\Gamma}(\bar{L}_5)$ will coincide with that of $\Gamma[\bar{g}_{\mu\nu}]$ using the definition (11). However, $\int \mathrm{d}^5x\,\sqrt{\bar{g}}$, while infinite, is not a constant but contains a factor of $\bar{L}_5$.)

To obtain an effective potential for $\bar{L}_5$ under adiabatic conditions, we need to find that Legendre transform of $\tilde{\Gamma}(\bar{L}_5)=\tilde{\Gamma}(\beta,\bar{L}_5)$ which is a function of S and $\bar{L}_5$, where S is the entropy. We make use of the following relations (J is a constant external cource):

$$Z(\beta,J)=\int[\mathrm{d}g_{\mu\nu}](\det\Delta)\exp\left(-S[g_{\mu\nu}]-\int \mathrm{d}^5x\,Jg_{55}(2\pi R_5)\right), \tag{40}$$

$$\frac{\partial}{\partial J}\ln Z|_{\beta}=-\bar{L}_5^2\beta\int \mathrm{d}^3\boldsymbol{x}\,, \tag{41}$$

$$S=\ln Z-\beta\frac{\partial}{\partial\beta}\ln Z|_J\,. \tag{42}$$

From (40)–(42) we find that

$$\frac{\partial}{\partial(\beta^{-1})}\left(\frac{\tilde{\Gamma}}{\beta}\right)\bigg|_{\bar{L}_5}=-\tilde{S}\,,\qquad \frac{\partial}{\partial(\bar{L}_5^2)}\left(\frac{\tilde{\Gamma}}{\beta}\right)\bigg|_{\beta}=-J\,. \tag{43a, b}$$

(The tilde above any extensive quantity denotes division by the infinite constant $\int d^3\boldsymbol{x}$, as in eq. (11).) The quantity we are looking for is therefore the energy per 3-volume,

$$\begin{aligned}\tilde{H} &= \beta^{-1}\tilde{\Gamma} + \beta^{-1}\tilde{S} \\ &= \beta^{-1}\tilde{\Gamma} + \beta\frac{\partial}{\partial\beta}\left(\frac{\tilde{\Gamma}}{\beta}\right),\end{aligned} \tag{44}$$

using (43a). $\tilde{H}$ satisfies

$$\left.\frac{\partial\tilde{H}}{\partial\tilde{S}}\right|_{\bar{L}_5} = \beta^{-1}, \qquad \left.\frac{\partial\tilde{H}}{\partial(\bar{L}_5^2)}\right|_{\tilde{S}} = -J. \tag{45a, b}$$

Eq. (45b) shows that, at fixed entropy, $\tilde{H}$ is minimized at the value of $\bar{L}_5$ corresponding to $J = 0$. This value of $\bar{L}_5$ is thus the physical equilibrium value, since J is merely a fiction inserted for mathematical convenience.

Using (43a), (44), and (35), we find that

$$\tilde{S} = \frac{15}{4\pi^2}\frac{1}{\bar{L}_5^3}\left[f\left(\frac{\beta}{\bar{L}_5}\right) - \frac{\beta}{\bar{L}_5}f'\left(\frac{\beta}{\bar{L}_5}\right)\right], \tag{46}$$

$$\tilde{H} = -\frac{15}{4\pi^2}\frac{1}{\bar{L}_5^4}f'\left(\frac{\beta}{\bar{L}_5}\right). \tag{47}$$

In the limits $\bar{L}_5 \ll \beta$ and $\bar{L}_5 \gg \beta$ we can see from eqs. (46), (36), and (37) that

$$\tilde{S} \propto \frac{\bar{L}_5}{\beta^4}, \qquad \bar{L}_5 \ll \beta, \tag{48a}$$

$$\tilde{S} \sim \frac{75\zeta(5)}{4\pi^2}\frac{\bar{L}_5}{\beta^4}, \qquad \bar{L}_5 \gg \beta. \tag{48b}$$

These limits may thus be expressed in the following equivalent ways:

$$\bar{L}_5 \overset{\ll}{\underset{\gg}{}} \beta \Leftrightarrow \tilde{S} \overset{\ll}{\underset{\gg}{}} \frac{1}{\bar{L}_5^3} \Leftrightarrow \tilde{S} \overset{\ll}{\underset{\gg}{}} \frac{1}{\beta^3}. \tag{49}$$

We now evaluate $\tilde{H}$ in the above limits. For $\bar{L}_5 \ll \beta$, $\tilde{H}$ is independent of β:

$$\tilde{H} \sim -\frac{15}{4\pi^2}\frac{1}{\bar{L}_5^4}. \tag{50}$$

For $\bar{L}_5 \gg \beta$, we can eliminate β using eq. (48b), obtaining

$$\tilde{H} \sim 4\left(\tfrac{1}{5}\tilde{S}\right)^{5/4}\left(\frac{4\pi^2}{15\zeta(5)\bar{L}_5}\right)^{1/4}. \tag{51}$$

So, $\tilde{H}(\bar{L}_5)$ at constant entropy has the qualitative features of the curve shown in fig. 2. We note that the minima still occur at $\bar{L}_5 = 0$ and $\bar{L}_5 = \infty$.

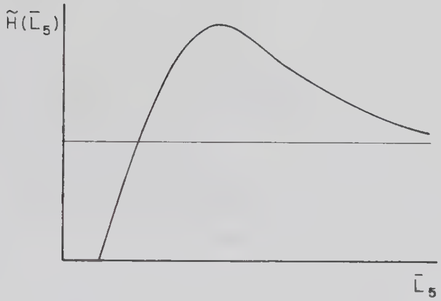


Fig. 2. Schematic representation of the effective potential at fixed entropy as a function of the circumference of the fifth dimension.

From fig. 2 and the inequalities (49), we see that expansion occurs when the amount of entropy (for a given value of $\bar{L}_5$ or β) is "large", and contraction when it is "small", a result which accords well with intuition.

Finally, we estimate the range of values of $\bar{L}_5$ and β for which we are justified in assuming that the average metric $\bar{g}_{\mu\nu}$ has the form (10), i.e. is flat. That *some* deviation from flatness must occur can be seen from the requirement that $\bar{g}_{\mu\nu}$ minimizes the effective action Γ. If we place no restrictions on the form of $\bar{g}_{\mu\nu}$, the "one-loop effective Einstein equation" $\delta\Gamma/\delta\bar{g}_{\mu\nu}(x)=0$ becomes, using eq. (3),

$$\begin{aligned} G^{\mu\nu}(x) &= 16\pi G_5 \frac{1}{\sqrt{\bar{g}}}\frac{\delta}{\delta\bar{g}_{\mu\nu}}[\tfrac{1}{2}\ln\det S_2 - \ln\det\Delta] \\ &\equiv 16\pi G_5 \frac{1}{\sqrt{\bar{g}}}\frac{\delta}{\delta\bar{g}_{\mu\nu}}\int \mathrm{d}^5x\,\gamma\,, \qquad (52)\end{aligned}$$

where $G^{\mu\nu}$ is the Einstein tensor formed from $\bar{g}_{\mu\nu}$. We see from (52) that the thermal and quantum fluctuations generate an effective stress-energy which will produce curvature.

To estimate the range of values of $\bar{L}_5$ and β for which we can ignore this curvature, we first estimate its magnitude using our previous flat-spacetime results. From (38), (39), (52) and $\int \mathrm{d}^5x = 2\pi R_5\beta\int \mathrm{d}^3\boldsymbol{x}$, we obtain

$$\gamma \sim -\frac{15\zeta(5)}{4\pi^2}\frac{1}{2\pi R_5\bar{L}_5^4}\,, \qquad \bar{L}_5 \ll \beta\,, \qquad (53a)$$

$$\gamma \sim -\frac{15\zeta(5)}{4\pi^2}\frac{\bar{L}_5}{2\pi R_5\beta^5}\,, \qquad \bar{L}_5 \gg \beta\,. \qquad (53b)$$

So, using

$$\frac{\partial\bar{L}_5}{\partial g_{55}} = \frac{1}{2}\frac{(2\pi R_5)^2}{\bar{L}_5}\,, \qquad \sqrt{\bar{g}} = \sqrt{\bar{g}_{55}}\,,$$

$$G^{55} = 16\pi G_5\left(\frac{15\zeta(5)}{2\pi^2}\right)\frac{(2\pi R_5)^2}{\bar{L}_5^7}\,, \qquad \bar{L}_5 \ll \beta\,, \qquad (54a)$$

$$G^{55} = 16\pi G_5\left(-\frac{15\zeta(5)}{8\pi^2}\right)\frac{(2\pi R_5)^2}{\bar{L}_5^2\beta^5}\,, \qquad \bar{L}_5 \gg \beta\,. \tag{54b}$$

As a quantitative measure of flatness we shall use the coordinate-invariant length scale $l \equiv |(2\pi R_5)^2/G^{55}|^{1/4}$. We are justified in regarding spacetime as flat if

$$l \gg \bar{L}_5\,. \tag{55}$$

Using (53a, b) the above condition becomes

$$\bar{L}_5^3\left[\frac{\pi}{120\zeta(5)}\right] \gg G_5\,, \qquad \bar{L}_5 \ll \beta\,, \tag{56a}$$

$$\beta^3\left(\frac{\beta}{\bar{L}_5}\right)^2\left[\frac{\pi}{30\zeta(5)}\right] \gg G_5\,, \qquad \bar{L}_5 \gg \beta\,. \tag{56b}$$

The Planck length L_P is defined as the square root of the four-dimensional Newton constant $G_4 = G_5/\bar{L}_5$. In terms of L_P, the "sufficient flatness condition" (55) becomes

$$\bar{L}_5 \gg 4\sqrt{\frac{15\zeta(5)}{2\pi}}\,L_P\,, \qquad \bar{L}_5 \ll \beta\,, \tag{57a}$$

$$\frac{\bar{L}_5}{\beta} \ll \left(\frac{\pi}{30\zeta(5)}\right)^{1/3}\left(\frac{\beta}{L_P}\right)^{2/3}\,, \qquad \bar{L}_5 \gg \beta\,. \tag{57b}$$

Thus, the results which we have obtained using a flat metric retain their validity only for values of the ratio $\bar{L}_5/\beta$ below an upper limit given by eq. (57b).

References

[1] Th. Kaluza, Sitz, Preuss. Akad. Wiss. Berlin; Math. Phys. Kl (1921) 966
O. Klein, Z. Phys. 37 (1926) 895
C.A. Orzalesi, Fortschr. der Phys. 29 (1981) 413

[2] P.G.O. Freund, Enrico Fermi Institute preprint EF1 82/83 (1982)

[3] T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141

[4] M.J. Duff, Nucl. Phys. B219 (1983) 389;
F. Englert, CERN preprint TH.3394;
E. Witten, Nucl. Phys. B186 (1981) 412
P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233

[5] E. Witten, Nucl. Phys. B195 (1982) 481

[6] (a) E.S. Abers and B.W. Lee, Phys. Reports 9C (1973) 1;
(b) C.W. Bernard, Phys. Rev. D9 (1974) 3312

[7] S.W. Hawking, Comm. Math. Phys. 55 (1977) 133

[8] Bateman Manuscript Project, Higher transcendental functions, vol. 3 (McGraw-Hill, New York, 1955) p. 195

[9] M. A. Rubin and B. D. Roth, Phys. Lett. 127B (1983)

CALCULATION OF GAUGE COUPLINGS AND COMPACT CIRCUMFERENCES FROM SELF-CONSISTENT DIMENSIONAL REDUCTION*

Philip CANDELAS and Steven WEINBERG

Center for theoretical Physics,
and
Theory Group, Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA

Received 19 May 1983
(Corrected version received 24 January 1984)

We consider a system of gravity plus free massless matter fields in $4+N$ dimensions, and look for solutions in which N dimensions form a compact curved manifold, with the energy-momentum tensor responsible for the curvature produced by quantum fluctuations in the matter fields. For manifolds of sufficient symmetry (including spheres, CP^N, and manifolds of simple Lie groups) the metric depends on only a single multiplicative parameter ρ^2, and the field equations reduce to an algebraic equation for ρ, involving the potential of the matter fields in the metric of the manifold. With a large number of species of matter fields, the manifold will be larger than the Planck length, and the potential can be calculated using just one-loop graphs. In odd dimensions these are finite, and give a potential of form C_N/ρ^4. Also there are induced Yang-Mills and Einstein–Hilbert terms in the effective 4-dimensional action, proportional to additional numerical coefficients, D_N and E_N. General formulas are given for the gauge coupling g^2 in terms of C_N and D_N, and the ratio $\rho^2/8\pi G$ in terms of C_N and E_N. Numerical values for C_N, D_N, and E_N are obtained for scalar and spinor fields on spheres of odd dimensionality N. It is found that the potential, g^2 and $\rho^2/8\pi G$ can all be positive but only when the compact manifold has $N=3+4k$ dimensions. (The positivity of the potential is needed for stability of the sphere against uniform dilations or contractions.) In this case, solutions exist either for spinor fields alone or for suitable mixes of spinor and scalar fields provided the ratio of the number of scalar fields to the number of fermion fields is not too large. Numerical values of the $\mathrm{O}(N+1)$ gauge couplings and $8\pi G/\rho^2$ are calculated for illustrative values of the numbers of spinor fields. It turns out that large numbers of matter fields are needed to make these parameters reasonably small.

1. Introduction

The calculation of the fine-structure constant has presented a challenge to theoretical physicists ever since the measurement of e and $\hbar$ in the early years of this century. We do not presume in this paper to offer a calculation of *the* fine-structure constant, $\frac{1}{137}$. Rather, we wish to show how fine-structure constants in general can be calculated in certain theories, in which the gauge fields arise from the metric of a higher-dimensional space.

---

* Research supported in part by the Robert A. Welch Foundation, the National Science Foundation under Grant PHY8205717 and the University of Texas Center for Theoretical Physics.

In the original Kaluza–Klein model [1], all electric charges are integer multiples of the ratio of $2\pi(16\pi G)^{1/2}$ and the circumference of the compact fifth dimension. However, this circumference is a free parameter of the model, one that is not determined even when we specify all parameters in the lagrangian. Thus, although this model accounts for the quantization of electric charge, it does not lead to a prediction of the basic unit of charge.

There are more general $(4+N)$-dimensional models [2, 3], in which N dimensions form a compact manifold, and a massless gauge field appears in four dimensions for each Killing vector of this manifold. A general prescription [4] has recently been given for calculating the various gauge couplings in such models in terms of the ratio of $2\pi(16\pi G)^{1/2}$ and various r.m.s. circumferences. Some of these models [3] exhibit a so-called "spontaneous compactification": the compact manifold is a solution of Einstein's field equations with non-zero curvature, with the energy-momentum tensor arising from monopole-like configurations of classical fields. Unfortunately, the "monopole" strength of these configurations is again a free parameter, unrelated to the parameters in the lagrangian. Hence, at least on a classical level, these models do not lead to a determination of the scale of the compact manifold, or of the gauge coupling constants.

In this paper we consider dynamical compactification of a different sort, proposed in ref. [4]. The $(4+N)$-dimensional space is again supposed to break up into a 4-dimensional Minkowski space and a curved compact N-dimensional manifold, with the curvature governed by Einstein's field equations. However, the energy-momentum tensor on the right-hand side of these equations is now supposed to arise not from some topologically non-trivial classical field configuration, but rather from the one-loop quantum fluctuations in various matter fields.

One naturally approaches any juxtaposition of quantum mechanics and general relativity with caution. This may explain why previous models of dynamical compactification have sought a solution in purely classical terms*. However, there is one class of models of quantum spontaneous compactification in which multi-loop gravitational corrections can be systematically neglected. These are the models outlined in ref. [4], with a large number f of light matter fields.

For an N-dimensional compact manifold whose linear dimensions are of order ρ, the one-loop energy density of f light matter fields is of order $f\rho^{-4-N}$. The $(4+N)$-dimensional gravitational constant $\bar{G}$ is of order $G\rho^{N}$, and the Einstein tensor

---

* Quantum matter fluctuations have been considered as a source of an energy-momentum tensor in higher-dimensional theories by the authors of ref. [5]. However, they consider the extra dimensions to form a circle or torus, so that there is no curvature to balance the energy-momentum tensor, and the field equations require that the matter potential itself be stationary. This is only possible for a mix of bosons and fermions, some of which must be massive, and the circumference of the compact manifold will depend on the values assumed for these masses. In contrast, in the present work the energy-momentum tensor is balanced by the curvature, and solutions are possible without mass parameters in the lagrangian, and with the scale of the compact manifold set by the gravitational constant.

$G_{\mu\nu}$ is of order ρ^{-2}. Hence the Einstein field equations give

$$\rho^{-2} \approx \bar{G} f \rho^{-4-N} \approx G f \rho^{-4},$$

and so

$$\rho^{2} \approx G f. \tag{1}$$

The L-loop gravitational corrections to the one-loop matter energy density are proportional to $f\bar{G}^{L-1}$, and hence on dimensional grounds are less for $L \geqslant 2$ than the one-loop matter terms by a factor

$$(\bar{G}\rho^{-N-2})^{L-1} \approx (G\rho^{-2})^{L-1} \approx (1/f)^{L-1}.$$

Also, the L-loop purely gravitational contributions to the energy density do not carry an over-all factor f, and are proportional to $\bar{G}^{L-1}$, so on dimensional grounds are less for $L \geqslant 1$ than the one-loop matter terms by a factor

$$\frac{1}{f}(\bar{G}\rho^{-N-2})^{L-1} \approx (1/f)^{L}.$$

Adding matter loops within virtual graviton lines adds equal numbers of additional $\bar{G}$ and f factors, and hence does not change the order in $1/f$.

We conclude that for f sufficiently large the scale of the compact manifold is of order $\sqrt{Gf}$, and can be calculated by including only one-loop matter contributions to the energy-momentum tensor*.

For an odd number N of compact dimensions, these models offer an added bonus. If we define our integrals by dimensional regularization then in the limit of 4 spacetime dimensions the one-loop energy density is actually finite [7]. (This is because dimensional regularization makes all divergences logarithmic, and the one-loop graphs are independent of dimensional couplings like $\bar{G}$, so any divergences would have to correspond to local counterterms with the dimensionality of the energy density in $4+N$ dimensions, i.e. $[\text{length}]^{-4-N}$. But the only local functions of the $(4+N)$-dimensional metric and its derivatives that are invariant under general coordinate transformations (and inversions) in the full space are those containing even numbers of derivatives, such as powers of the curvature scalar, and none of these can have the odd dimensionality $4+N$). The general formal considerations of this paper will apply to compact manifolds of arbitrary dimensionality N, but our numerical calculations will be limited to the case of odd N.

---

* The approximation used in our paper, based on the classical Einstein field equations with an energy-momentum tensor generated by one-loop quantum fluctuations in a large number of matter fields had previously been employed by Hartle and Horowitz [6]. Their work dealt only with 4-dimensional spacetimes.

2. Solution of the field equations

We consider a $(4+N)$-dimensional theory of gravity plus a number of massless matter fields. The Einstein field equations are*

$$\bar{R}^{NM} - \tfrac{1}{2}\bar{g}^{NM}(\bar{R}+\bar{\Lambda}) = -8\pi\bar{G}\bar{T}^{NM}\,, \tag{2}$$

where $\bar{g}_{NM}$ is the $(4+N)$-dimensional metric; $\bar{R}^{NM}$ and $\bar{R}$ are the Ricci tensor and curvature scalar calculated from $\bar{g}_{NM}$; $\bar{G}$ and $\bar{\Lambda}$ are the gravitational and cosmological constants in $4+N$ dimensions; and $\bar{T}^{NM}$ is the $(4+N)$-dimensional energy-momentum tensor, given as usual by

$$\tfrac{1}{2}\sqrt{\bar{g}}\,\bar{T}^{NM} = \delta\bar{I}/\delta\bar{g}_{NM}\,, \tag{3}$$

with $\bar{I}$ the effective matter action, produced here by one-loop quantum fluctuations.

We seek a vacuum solution with Poincaré invariance in 4 dimensions. That is, the metric is taken to have the components

$$\bar{g}_{\mu\nu} = \eta_{\mu\nu}\,, \qquad \bar{g}_{\mu n} = 0\,, \qquad \bar{g}_{nm} = \tilde{g}_{nm}(y)\,, \tag{4}$$

where $\eta_{\mu\nu}$ is the usual Minkowski metric for spacetime coordinates x^{μ}, and $\tilde{g}_{nm}(y)$ is the metric of some compact manifold with coordinates y^{m}. The Ricci tensor and curvature scalar appearing in (2) are then

$$\bar{R}^{\mu\nu} = \bar{R}^{\mu n} = \bar{R}^{n\mu} = 0\,, \tag{5}$$

$$\bar{R}^{nm} = \tilde{R}^{nm}(y)\,, \tag{6}$$

$$\bar{R} = \tilde{R}(y)\,, \tag{7}$$

where $\tilde{R}^{nm}$ and $\tilde{R}$ are the N-dimensional Ricci tensor and curvature scalar calculated from $\tilde{g}_{nm}(y)$. Also, the vacuum Poincaré invariance gives the energy-momentum tensor the structure

$$\bar{T}^{\mu\nu} = B(y)\eta^{\mu\nu}\,, \tag{8}$$

$$\bar{T}^{\mu n} = \bar{T}^{n\mu} = 0\,, \tag{9}$$

$$\bar{T}^{nm} = \tilde{T}^{nm}(y)\,. \tag{10}$$

The field equations (2) thus reduce to

$$\tfrac{1}{2}(\tilde{R}(y)+\bar{\Lambda}) = -8\pi\bar{G}B(y)\,, \tag{11}$$

$$\tilde{R}^{nm}(y) - \tfrac{1}{2}\tilde{g}^{nm}(y)[\tilde{R}(y)+\bar{\Lambda}] = -8\pi\bar{G}\tilde{T}^{nm}(y)\,. \tag{12}$$

In calculating $\tilde{T}^{nm}(y)$, we must evaluate a variational derivative with respect to $\bar{g}_{nm}$, so we can fix $\bar{g}_{\mu n}$ and $\bar{g}_{\mu\nu}$ at the values given by (4). The matter action is then

$$\bar{I} = -\int \mathrm{d}^4x\, V[\bar{g}_{nm}]\,, \tag{13}$$

---

* We must include a cosmological constant in $4+N$ dimensions and tune its value to ensure that the field equations have a solution with a flat 4-dimensional Minkowski space. This of course is an unpleasant feature of our theory, and all other existing theories.

with V the "potential" for the classical field $\tilde{g}_{nm}(x, y)$. The nm energy-momentum components are given by eq. (3) for $\tilde{g}_{nm}(x, y) = \tilde{g}_{nm}(y)$ as

$$\tfrac{1}{2}\sqrt{\tilde{g}(y)}\,\tilde{T}^{nm}(y) = -\frac{\delta V[\tilde{g}]}{\delta \tilde{g}_{nm}(y)}. \tag{14}$$

We will return later to the calculation of the other component, $B(y)$.

The nm field equations (12) are in general not easy to solve. However, as discussed in appendix A, there is a certain class of manifolds whose symmetries dictate (i) that $\tilde{g}_{nm}(y)$ is fixed up to an over-all factor ρ^2, and (ii) all other tensors like $\tilde{T}_{nm}(y)$ and $\tilde{R}_{nm}(y)$ formed by variational derivatives of scalars with respect to $\tilde{g}_{nm}(y)$ are proportional to $\tilde{g}_{nm}(y)$, with constant coefficients. These manifolds consist of the homogeneous coset spaces G/H for which generators of G that are not in H form a representation of H that is either real and irreducible or the direct sum of a complex irreducible representation and its complex conjugate. (This includes spheres, CP^N, and manifolds of simple groups.) For manifolds of this class, we can normalize the one free parameter ρ^2 in the metric by specifying that

$$\tilde{R}_{nm}(y) = -(N-1)\rho^{-2}\tilde{g}_{nm}(y)\,. \tag{15}$$

(The factor $-(N-1)$ is inserted here so that for spheres ρ will simply be the radius.) Also, the energy-momentum tensor here takes the form

$$\tilde{T}_{nm}(y) = A\tilde{g}_{nm}(y)\,, \tag{16}$$

and homogeneity implies that B is y independent.

We can calculate A by contracting (14) with $\tilde{g}_{nm}(y)$ and integrating over the manifold:

$$-\tfrac{1}{2}AN\Omega_N = \int \mathrm{d}^N y\,\tilde{g}_{nm}(y)\,\frac{\delta V[\tilde{g}]}{\delta \tilde{g}_{nm}(y)}\,, \tag{17}$$

where Ω_N is the volume of the manifold

$$\Omega_N \equiv \int \mathrm{d}^N y\,\sqrt{\tilde{g}(y)}\,. \tag{18}$$

The right-hand side of (17) is just the logarithmic rate of change of $V[\tilde{g}]$ when we vary the over-all scale of $\tilde{g}_{nm}(y)$, so we can write this as

$$-\tfrac{1}{2}N\Omega_N A = \rho^2\frac{\mathrm{d}V}{\mathrm{d}\rho^2}\,. \tag{19}$$

To calculate B, it is only necessary to notice that if we had used a spacetime metric with $g_{\mu\nu} = \lambda^2\eta_{\mu\nu}$, then a factor $\sqrt{g} = \lambda^4$ would have appeared in (13), so

$$\tfrac{1}{2}\bar{T}^{\mu}{}_{\mu} = \left(g_{\mu\nu}\frac{\delta \bar{I}}{\delta g_{\mu\nu}}\right)_{\lambda=1} = -\frac{V}{\Omega_N}\left(\lambda^2\frac{d}{d\lambda^2}\lambda^4\right)_{\lambda=1} = -\frac{2V}{\Omega_N}\,,$$

and therefore

$$\Omega_N B = -V. \tag{20}$$

Note that this gives the one-loop energy density $\bar{T}^{00}$ per unit volume in $(4+N)$-dimensional space as $-B = V/\Omega_N$, in accord with the usual interpretation of V as the energy density in 4-dimensional space. This result would apply to any homogeneous manifold, not only those with the irreducibility property discussed in appendix A.

The factors Ω_N in the formulas (19) and (20) for A and B will be cancelled by the Ω_N factor in the well-known relation between the gravitational constants G_0 and $\bar{G}$ in 4 and $4+N$ dimensions

$$\bar{G} = G_0 \Omega_N. \tag{21}$$

Using (15), (16), (19), (20), and (21), the field equations (11) and (12) now become

$$-\frac{1}{2}\left[\frac{-N(N-1)}{\rho^2} + \bar{\Lambda}\right] = 8\pi G_0 V, \tag{22}$$

$$\frac{-(N-1)}{\rho^2} - \frac{1}{2}\left[\frac{-N(N-1)}{\rho^2} + \bar{\Lambda}\right] = \frac{8\pi G_0}{N}\rho\frac{\mathrm{d}V}{\mathrm{d}\rho}. \tag{23}$$

Using the first equation to determine $\bar{\Lambda}$, the second becomes

$$\frac{N-1}{8\pi G_0 \rho^2} = V(\rho) - \frac{\rho}{N}\frac{\mathrm{d}V(\rho)}{\mathrm{d}\rho}. \tag{24}$$

Eq. (24) can be solved for the radius parameter ρ, and given the symmetries of the manifold, this completely determines its metric.

We now make the further assumptions that the matter fields are massless in $4+N$ dimensions and that N is odd. Otherwise the radius we calculate from (24) would depend on the mass parameters in the lagrangian, and there would be no hope of obtaining a definite value for the scale of gauge couplings without further information. Since there is no one-loop conformal anomaly in odd dimensions [7], naive dimensional analysis tells us that the potential here has the form

$$V = \frac{C_N}{\rho^4}, \tag{25}$$

with C_N a pure number, depending only on the symmetries of the manifold, on N, and on the numbers of massless fields of various types. Eq. (24) gives immediately*

$$\rho^2 = \frac{8\pi G_0 (N+4) C_N}{N(N-1)}, \tag{26}$$

---

* Note that the value for the radius ρ given by eq. (26) blows up for $N = 1$. This just reflects the fact that the compact one-dimensional manifold S^1 is flat, so there is no solution of the field equations in the original Kaluza–Klein model with a non-vanishing energy-momentum tensor.

while (22) gives the cosmological constant

$$\bar{\Lambda}=\frac{N^2(N-1)^2(N+2)}{8\pi G_0(N+4)^2C_N}\,. \tag{27}$$

As we will see in sect. 7, the stability of the compact manifold requires that G_0 be positive. This being the case, a solution can be found for ρ if and only if C_N is *positive*, and in this case $\bar{\Lambda}$ must be tuned to a definite positive value to obtain a flat spacetime.

3. Calculation of the potential

We consider here the potential V due to b real massless scalar fields and f Dirac massless spin-$\frac{1}{2}$ fields in $4+N$ dimensions. Each of these fields will appear in 4 dimensions as an infinite tower of spin-0 and spin-$\frac{1}{2}$ particles with masses given by the eigenvalues M_l^2 and m_l of the N-dimensional laplacian and Dirac differential operators. The potential produced by a single real spin-0 particle of mass M is known to be [8]

$$V_0(M)=\frac{-i}{2(2\pi)^4}\int \mathrm{d}^4k\,\ln(k^2+M^2-i\varepsilon)\,,$$

while for spin-$\frac{1}{2}$, we just replace M with m and multiply by -4. The total potential is then

$$V=b\sum_l D_lV_0(M_l)-4f\sum_l d_lV_0(m_l)\,, \tag{28}$$

where D_l and d_l are the degeneracies of the eigenvalues M_l and m_l of the laplacian and Dirac operators.

We will use dimensional regularization to evaluate $V_0(M)$. An elementary calculation gives

$$V_0(M)=-\tfrac{1}{2}(4\pi)^{-n/2}\Gamma(-\tfrac{1}{2}n)M^n\,, \tag{29}$$

where n is the spacetime dimensionality, later to be set equal to 4. For manifolds whose metric is characterized by a single over-all scale factor ρ^2 such as those discussed in appendix A, the eigenvalues are

$$M_l=\Lambda_l/\rho\,, \tag{30}$$

$$m_l=\lambda_l/\rho \tag{31}$$

where Λ_l and λ_l as well as the degeneracies D_l and d_l are ρ independent, depending only on the symmetries and dimensionality of the compact manifold. For instance, for massless scalars on a sphere S^N we have [9]*

$$\Lambda_l^2=l(l+N-1)\,,$$

$$D_l=\frac{(2l+N-1)(l+N-2)!}{(N-1)!\,l!}\,.$$

---

* Our conventions regarding standard functions follow those of this reference.

Using (29)–(31) in (28), the potential is

$$V=\lim_{n\to 4}\left[-\tfrac{1}{2}(4\pi)^{-n/2}\Gamma(-\tfrac{1}{2}n)\rho^{-n}\{b\zeta_N^{(0)}(-n)-4f\zeta_N^{(1/2)}(-n)\}\right], \tag{32}$$

where the ζ_N are generalized zeta functions

$$\zeta_N^{(0)}(z)=\sum_l D_l\Lambda_l^{-z}, \tag{33}$$

$$\zeta_N^{(1/2)}(z)=\sum_l d_l\lambda_l^{-z}. \tag{34}$$

All terms in the power series for $\Gamma(-\frac{1}{2}n)$ and $\zeta_N(-n)$ are positive, so at first sight it looks like the condition $C_N>0$ found necessary in sect. 2 can be satisfied if and only if there is a sufficiently large ratio of fermion to boson field types. However, both of these series diverge for $n>0$, and must therefore be defined by analytic continuation for $n\to 4$. Only then can we tell the signs.

The analytic continuation is carried out by two different methods in appendices B, D in the special case where the compact manifold is a sphere S^N of odd dimensionality N. The results confirm the general expectation that there should be no one-loop ultraviolet divergences in odd dimensions: the generalized zeta functions $\zeta_N^{(0)}(z)$ and $\zeta_N^{(1/2)}(z)$ are found like the ordinary Riemann zeta function $\zeta(z)$ to have a simple zero at negative even integer arguments, which cancels the pole in $\Gamma(-\frac{1}{2}n)$ as $n\to 4$. From (32) we see then that $V=C_N/\rho^4$, with

$$C_N=bC_N^{(0)}+fC_N^{(1/2)}, \tag{35}$$

$$C_N^{(0)}=-\frac{1}{32\pi^2}\zeta_N^{(0)\prime}(-4), \tag{36}$$

$$C_N^{(1/2)}=\frac{1}{8\pi^2}\zeta_N^{(1/2)\prime}(-4). \tag{37}$$

The most convenient way of calculating C_N numerically is provided by the integral representations developed in appendix B. For spheres and for the case of minimally coupled scalar fields these yield

$$\begin{aligned}C_N^{(0)}=\frac{1}{4\pi\nu}\int_0^{\pi/2}\frac{\mathrm{d}\theta}{[2\cosh(\frac{1}{2}\pi\tan\theta)]^{2\nu}}&\left\{\left(\frac{\nu^3}{\pi^3}\cos\theta\sin 3\theta-\frac{15\nu}{\pi^5}\cos^3\theta\sin 5\theta\right)\right.\\&\times\sinh(\nu\pi\tan\theta)+\left(\frac{6\nu^2}{\pi^4}\cos^2\theta\cos 4\theta-\frac{15}{\pi^6}\cos^4\theta\cos 6\theta\right)\\&\left.\times\cosh(\nu\pi\tan\theta)\right\},\end{aligned} \tag{38}$$

while for fermions we find

$$C_N^{1/2}=\frac{(-1)^{\nu+1}3\cdot 2^{\nu+1}}{\pi^6}\int_0^{\pi/2}\frac{\mathrm{d}\theta\cos^3\theta\cos 5\theta}{[2\cosh(\frac{1}{2}\pi\tan\theta)]^{2\nu+1}}, \tag{39}$$

TABLE 1

Values of $C_N^{(0)}$, the one-loop potential for a single massless real spin-0 field on an N-dimensional sphere of unit radius

N	$C_N^{(0)}$
1 (a)	-5.05576×10^{-5}
(b)	4.73982×10^{-5}
3	7.56870×10^{-5}
5	4.28304×10^{-4}
7	8.15883×10^{-4}
9	1.13389×10^{-3}
11	1.32932×10^{-3}
13	1.37403×10^{-3}
15	1.25249×10^{-3}
17	9.55916×10^{-4}
19	4.79352×10^{-4}
21	-1.79909×10^{-4}

Since the circle S^1 is multiply connected, for $N=1$ we have the choice of whether or not to allow the field to change sign around the circle. These two cases of untwisted and twisted fields are labelled here as (a) and (b), respectively.

TABLE 2

Values of $C_N^{(1/2)}$, the one-loop potential for a single massless Dirac spin-$\frac{1}{2}$-field on an N-dimensional sphere of unit radius

N	$C_N^{(1/2)}$
1 (a)	$+2.022304\times10^{-4}$
(b)	-1.895928×10^{-4}
3	$+1.945058\times10^{-4}$
5	-1.140405×10^{-4}
7	$+5.958744\times10^{-5}$
9	-2.99172×10^{-5}
11	$+1.477709\times10^{-5}$
13	-7.242740×10^{-6}
15	$+3.537614\times10^{-6}$
17	-1.725405×10^{-6}
19	$+8.412070\times10^{-7}$
21	-4.101970×10^{-7}

As in the scalar case (a) and (b) denote the values appropriate to the field being, respectively, untwisted and twisted.

where

$$\nu \equiv \tfrac{1}{2}(N-1)\,. \tag{40}$$

Numerical results are presented in tables 1 and 2 (the corresponding results for conformally coupled scalar fields are presented in table 3). The asymptotic values of the $C_N^{(j)}$ for large N are derived in appendix C:

$$C_N^{(0)} = -\frac{\nu^2}{16\pi^2\mu^2}\left(1+\mathrm{O}\left(\frac{1}{\mu}\right)\right), \tag{41}$$

where, in this context, μ is the root of

$$\mathrm{e}^{\mu}/\mu = \tfrac{2}{3}\nu\,, \tag{42}$$

$$C_N^{(1/2)} = \frac{(-)^{\nu+1}\cdot 3}{2^{\nu}\pi^{13/2}(\nu+\frac{1}{2})^{1/2}}\left\{1-\left(\frac{30}{\pi^2}-\frac{1}{8}\right)\frac{1}{(\nu+\frac{1}{2})}+\mathrm{O}\left(\frac{1}{(\nu+\frac{1}{2})^2}\right)\right\}. \tag{43}$$

We believe that the sign change in $C_N^{(0)}$ that occurs between $N=19$ and $N=21$ is the one "predicted" by the asymptotic form (41) and we therefore expect that the sign of $C_N^{(0)}$ will remain negative for all $N>21$. On the other hand for fermions the alternation of sign, expected asymptotically for large-N, sets in immediately at $N=3$, and this alternation will continue for all N.

We see that a solution can be found (i.e. $C_N>0$) with any numbers of massless scalars and fermions only for spaces of total dimensionality $N+4$ equal to 7, 11, 15, 19, and 23. There is no solution at all if the total dimensionality is 25, 29, 33, $\cdots$. A solution can be found provided the ratio of fermions to bosons is sufficiently large for $N+4=27, 31, 35, \cdots$ and provided the ratio of bosons to fermions is sufficiently large for $N+4=9, 13, 17$, and 21. However, as we shall see, there are additional positivity conditions that exclude most of these possibilities.

TABLE 3

Values of $C_N^{\text{conformal}}$, the one-loop potential for a single massless conformally coupled scalar field on an N-dimensional sphere of unit radius

N	$C_N^{\text{conformal}}$
3	$+7.145890\times10^{-6}$
5	-7.857102×10^{-7}
7	$+7.049427\times10^{-8}$
9	-1.821234×10^{-9}
11	-1.566188×10^{-9}
13	$+6.501802\times10^{-10}$
15	-1.943404×10^{-10}
17	$+5.201765\times10^{-11}$
19	-1.322957×10^{-11}
21	$+3.274453\times10^{-12}$

4. Gauge and gravitational couplings

When a $(4+N)$-dimensional space breaks up into a 4-dimensional Minkowski space and a compact manifold, the perturbations of this metric appear in 4 dimensions in part as a set of massless fields: the Yang–Mills fields $A^{\mu}_{\alpha}(x)$ and the gravitational field $g_{\mu\nu}(x)$. In ref. [4] it was shown that the classical Einstein–Hilbert action of pure gravity in $4+N$ dimensions yields in 4 dimensions an action for these fields of the form:

$$\begin{aligned}\bar{I} &\equiv -\frac{1}{16\pi\bar{G}}\int \bar{R}\sqrt{\bar{g}}\,\mathrm{d}^{4+N}\bar{x}\\ &= -\frac{1}{16\pi G_0}\left[\int R\sqrt{g}\,\mathrm{d}^4x+\frac{1}{4}\sum_e \frac{g_e^2N_e^2s_e^2}{4\pi^2}\int F_e^{\mu\nu}F_{e\mu\nu}\sqrt{g}\,\mathrm{d}^4x+\cdots\right],\end{aligned}\tag{44}$$

where in our present notation $F_e^{\mu\nu}$ are the Yang–Mills curls of those gauge fields A_e^{μ} that correspond to closed Killing curves of the compact manifold; s_e is the r.m.s. circumference of the manifold along these curves; g_e is the corresponding gauge coupling, defined as the lowest eigenvalue of the generator t_e in the faithful N_e-valued representation whose structure constants $C_{ef}{}^{g}$ appear in $F_e^{\mu\nu}$; and (as in sect. 3)

$$G_0 \equiv \bar{G}\Big/\int \mathrm{d}^Ny\sqrt{\bar{g}}(y)\,.$$

If this were the whole story, then G_0 could be identified as the Newton gravitational coupling constant G, and the normalization condition for the Yang–Mills fields would yield the formula for g_e derived in ref. [4]:

$$g_e=\frac{2\pi(16\pi G)^{1/2}}{N_es_e}\,.$$

However, we are now assuming that there are many species of matter fields, so that one-loop radiative corrections can be important[*] despite the smallness of G/ρ^2.

In general, radiative corrections generate induced Einstein–Hilbert and Yang–Mills terms in the effective four-dimensional action, of the form

$$I_{\text{induced}}=-\frac{1}{4}\sum_e D_{Ne}g_e^2\int F_e^{\mu\nu}F_{e\mu\nu}\sqrt{g}\,\mathrm{d}^4x-\frac{E_N}{\rho^2}\int R\sqrt{g}\,\mathrm{d}^4x+\cdots\,.\tag{45}$$

The D_{Ne} and E_N are new dimensionless coefficients, which for massless matter fields on one-parameter manifolds (see appendix A) of odd dimensionality are (like C_N) ρ-independent numerical quantities roughly of the order of the number of species of matter fields. Combining (44) and (45), we see that the true Newton constant G is given by

$$\frac{1}{16\pi G}=\frac{1}{16\pi G_0}+\frac{E_N}{\rho^2}\,,\tag{46}$$

---

[*] The induced Einstein–Hilbert and Yang–Mills terms are calculated in the original Kaluza–Klein model with $N=1$ by Toms, ref. [10]. Toms' preprint was instrumental in reminding us of the importance of these terms.

and the normalization condition for the gauge fields is

$$1=\frac{g_e^2 N_e^2 s_e^2}{4\pi^2\cdot 16\pi G_0}+D_{Ne}g_e^2\,. \tag{47}$$

In particular, for spheres S^N, all gauge couplings and D-coefficients have common values g and D_N; the r.m.s. circumference is

$$s_e=2\pi\rho\sqrt{\frac{2}{N+1}}\,,$$

and $N_e=1$; so (47) here reads:

$$1=g^2\left[\frac{\rho^2}{16\pi G_0}\frac{2}{N+1}+D_N\right]. \tag{48}$$

The previous results of ref. [4] gave ρ^2/G and $1/g^2$ of the order of the number of species of matter fields, and hence of the order of E_N and D_N, so the radiative corrections represented by E_N and D_N in (46) and (47)–(48), are of the same order as the other, classical, terms. However higher radiative corrections are suppressed by factors $1/f$ and $1/b$.

Now let us apply what we have previously learned about the size of the compact manifold. Recall that eq. (26) gives ρ^2 in terms of G_0, not G:

$$\rho^2=\frac{8\pi G_0(N+4)C_N}{N(N-1)}\,, \tag{49}$$

so eq. (48) gives the O$(N+1)$ coupling as

$$g^2=\left[\frac{(N+4)C_N}{N(N^2-1)}+D_N\right]^{-1}. \tag{50}$$

This is what we wanted: a formula for the gauge coupling that depends only on the dimensionality and shape of the compact manifold and the numbers of various types of massless matter fields, and that gives a small gauge coupling if these numbers are large. However, we now encounter a new positivity condition

$$\frac{(N+4)C_N}{N(N^2-1)}+D_N>0\,. \tag{51}$$

If this is not satisfied, then the Yang–Mills term in the effective $(4+N)$-dimensional action has the "wrong" sign, and the vacuum is unstable with regard to exponentially growing Yang–Mills waves.

Similarly, we can use (49) to eliminate G_0 in eq. (46), and solve for the squared radius ρ^2 in terms of the observed gravitational constant G:

$$\frac{\rho^2}{8\pi G}=\frac{(N+4)C_N}{N(N-1)}+2E_N\,. \tag{52}$$

Thus as before the size of the compact manifold is larger than the Planck length if there are many matter fields. The immediate consequence of (52) is another positivity condition

$$\frac{(N+4)C_N}{N(N-1)}+2E_N>0\,. \tag{53}$$

If this is not satisfied, then the Einstein–Hilbert term in the effective action has the "wrong" sign, and the vacuum is unstable with regard to exponentially growing gravitational waves.

5. Calculation of D_N and E_N

We now turn to a calculation of the new coefficients D_N and E_N. We note first that in four dimensions the one-loop graphs involving a single minimally coupled particle of mass M and gauge coupling $g_e Q_e$ will produce the induced Einstein–Hilbert and Yang–Mills terms*

$$-\tfrac{1}{4}\sum_e \mathscr{D}_e(M)g_e^2 \int F_e^{\mu\nu}F_{e\mu\nu}\sqrt{g}\,\mathrm{d}^4x$$

$$-\mathscr{E}(M)\int R\sqrt{g}\,\mathrm{d}^4x\,,$$

where for scalars

$$\mathscr{D}_e(M)=\frac{1}{96\pi^2}\lim_{n\to4}\Gamma(2-\tfrac{1}{2}n)M^{n-4}Q_e^2\,, \tag{54}$$

$$\mathscr{E}(M)=\frac{1}{192\pi^2}\lim_{n\to4}\Gamma(1-\tfrac{1}{2}n)M^{n-2}\,. \tag{55}$$

If in $4+N$ dimensions we have b massless scalars, then in 4 dimensions we will observe b infinite towers of massive scalars: say bD_l scalars of mass M_l. Thus the D_N and E_N coefficients for each species will be given by

$$D_N^{(0)\prime}=\frac{1}{96\pi^2}\lim_{n\to4}\Gamma\left(2-\frac{n}{2}\right)\sum_l D_l M_l^{n-4}Q_{le}^2\,, \tag{56}$$

$$E_N^{(0)}=\frac{1}{192\pi^2}\lim_{n\to4}\Gamma\left(1-\frac{n}{2}\right)\sum_l D_l M_l^{n-2}\rho^2\,, \tag{57}$$

where the Q_{le}^2 are the mean square charges (in units of g_e^2) of the scalars of mass M_l.

For spinors induced gravity and Yang–Mills terms arise similarly although the situation with regard to the induced Yang–Mills term is complicated by the fact that, through the process of dimensional reduction, the Dirac equation acquires a non-minimal coupling to the background gauge field. Fortunately these effects have been considered by Awada and Toms [10] who find that for each species

---

* The contributions of individual massive particles to the induced Einstein–Hilbert and Yang–Mills actions in 4 dimensions can be inferred from the work of deWitt, Toms, and Awada and Toms [10], or taken for spin-0 (for E_N) from S. Weinberg, ref. [11] and for spin-$\frac{1}{2}$ from Zee, ref. [12].

$$D_N^{(1/2)} = -\frac{1}{192\pi^2}\frac{(20\nu^2+28\nu+5)}{(\nu+1)^2}\lim_{n\to4}\Gamma\left(2-\frac{n}{2}\right)\sum_l d_l m_l^{n-4} q_{le}^2\,, \tag{58}$$

$$E_N^{(1/2)} = \frac{1}{96\pi^2}\lim_{n\to4}\Gamma\left(1-\frac{n}{2}\right)\sum_l d_l m_l^{n-2}\rho^2\,, \tag{59}$$

where d_l and q_{le}^2 are the number and mean square charges of spinors of mass m_l and

$$\nu \equiv \tfrac{1}{2}(N-1)\,.$$

Now let us specialize to the case where the compact manifold is a sphere S^N of radius ρ. As we saw in the calculation of C_N, the masses and degeneracies for spin-0 are

$$M_l^2 = l(l+N-1)/\rho^2\,, \tag{60}$$

$$D_l = \frac{(2l+N-1)(l+N-2)!}{(N-1)!\,l!}\,, \tag{61}$$

and for spin-$\frac{1}{2}$

$$m_l = (l+\tfrac{1}{2}N)/\rho\,, \tag{62}$$

$$d_l = \frac{2^{(N+1)/2}(l+N-1)!}{(N-1)!\,l!}\,. \tag{63}$$

The mean square charges Q_{le}^2 and q_{le}^2 are the same for all $O(N+1)$ generators; they are calculated in appendix E as

$$Q_l^2 = [l(l+N-1)]/\tfrac{1}{2}N(N+1)\,, \tag{64}$$

$$q_l^2 = [l^2+lN+N(N+1)/8]/\tfrac{1}{2}N(N+1)\,. \tag{65}$$

Using methods like those previously applied to the calculation of C_N, we show in appendix F that eqs. (60)–(65) allow us to write (56)–(59) as the integrals:

$$D_N^{(0)} = -\frac{4}{N(N+1)}E_N^{(0)}\,, \tag{66}$$

$$\begin{aligned} E_N^{(0)} = \frac{1}{48\pi^4\nu}\int_0^{\pi/2} & d\theta\,[2\cosh(\tfrac{1}{2}\pi\tan\theta)]^{-2\nu} \\ & \times\{3\nu\cos\theta\sin3\theta\sinh(\nu\pi\tan\theta) \\ & +\left(\frac{3}{\pi}\cos^2\theta\cos4\theta-\nu^2\pi\cos2\theta\right)\cosh(\nu\pi\tan\theta)\}\,, \end{aligned} \tag{67}$$

$$\begin{aligned} D_N^{(1/2)} = \frac{(20\nu^2+28\nu+5)}{16(\nu+1)^2} & \left\{\frac{16}{N(N+1)}E_N^{(1/2)}\right. \\ & \left.+\left(\frac{N-1}{N+1}\right)\frac{(-1)^{\nu+1}2^\nu}{12\pi^2}\int_0^{\pi/2} d\theta[2\cosh(\tfrac{1}{2}\pi\tan\theta)]^{-2\nu-1}\right\}\,, \end{aligned} \tag{68}$$

$$E_N^{(1/2)} = \frac{(-1)^{\nu+1}2^{\nu}}{12\pi^2} \int_0^{\pi/2} d\theta [2\cosh(\tfrac{1}{2}\pi\tan\theta)]^{-2\nu-1} \cos\theta \cos 3\theta . \qquad (69)$$

where as before, $\nu \equiv \frac{1}{2}(N-1)$.

Instead of giving the values of D_N and E_N, we present in tables 4 and 5 the values of the combinations

$$\beta_N^{(j)} \equiv \frac{(N+4)}{N(N-1)} C_N^{(j)} + 2E_N^{(j)} , \qquad (70)$$

$$\gamma_N^{(j)} \equiv \frac{(N+4)}{N(N^2-1)} C_N^{(j)} + D_N^{(j)} , \qquad (71)$$

whch appear in the formulas for ρ^2/G and g^2.

6. Results

For spherical compact manifolds, the radius ρ and $O(N+1)$ coupling g are given by eqs. (52) and (50) as

$$\rho^2 = 8\pi G \beta_N , \qquad (72)$$

$$g^2 = 1/\gamma_N , \qquad (73)$$

where according to (54) and (55)

$$\beta_N = b\beta_N^{(0)} + f\beta_N^{(1/2)} , \qquad (74)$$

$$\gamma_N = b\gamma_N^{(0)} + f\gamma_N^{(1/2)} , \qquad (75)$$

with $\beta_N^{(j)}$ and $\gamma_N^{(j)}$ the coefficients (70) and (71). As mentioned earlier it is crucial that β_N and γ_N as well as C_N be positive. A glance at tables 4 and 5 shows that for $N = 5 \pmod 4$, both $\gamma_N^{(0)}$ and $\gamma_N^{(1/2)}$ are negative, so in this case there is no way

TABLE 4

Values of $\beta_N^{(0)}$ and $\gamma_N^{(0)}$

N	$\beta_N^{(0)}$	$\gamma_N^{(0)}$
3	3.9281071×10^{-4}	$-2.8676136 \times 10^{-5}$
5	7.8068082×10^{-4}	$-7.0734863 \times 10^{-6}$
7	1.1159701×10^{-3}	$-5.5139621 \times 10^{-6}$
9	1.4314887×10^{-3}	$-6.7881832 \times 10^{-6}$
11	1.7353643×10^{-3}	$-8.4410152 \times 10^{-6}$
13	2.0310803×10^{-3}	$-9.9788801 \times 10^{-6}$
15	2.3205490×10^{-3}	$-1.1311056 \times 10^{-5}$
17	2.6049752×10^{-3}	$-1.2443484 \times 10^{-5}$
19	2.8851865×10^{-3}	$-1.3403666 \times 10^{-5}$
21	3.1617857×10^{-3}	$-1.4220508 \times 10^{-5}$

TABLE 5

Values of $\beta_N^{(1/2)}$ and $\gamma_N^{(1/2)}$

N	$\beta_N^{(1/2)}$	$\gamma_N^{(1/2)}$
3	$+3.3832530\times10^{-4}$	$+5.0098703\times10^{-4}$
5	-1.0389980×10^{-4}	-2.6068446×10^{-4}
7	$+4.0049456\times10^{-5}$	$+1.2830668\times10^{-4}$
9	-1.6805816×10^{-5}	-6.1937531×10^{-5}
11	$+7.3681781\times10^{-6}$	$+2.9734556\times10^{-5}$
13	-3.3174183×10^{-6}	-1.4265962×10^{-5}
15	$+1.5203817\times10^{-6}$	$+6.8526826\times10^{-6}$
17	-7.0568431×10^{-7}	-3.2977732×10^{-6}
19	$+3.3066589\times10^{-7}$	$+1.5902169\times10^{-6}$
21	-1.5608669×10^{-7}	-7.6834453×10^{-7}

to avoid a negative value for g^2, which would imply a vacuum that is not even a local minimum of the energy. We therefore must restrict our attention to the cases $N=3 \pmod 4$. For these cases, all three of $C_N^{(1/2)}$, $\beta_N^{(1/2)}$ and $\gamma_N^{(1/2)}$ are positive so fermions alone furnish a solution. If $N=3 \pmod 4$ and $N<21$, then $C_N^{(0)}$ and $\beta_N^{(0)}$ are positive while $\gamma_N^{(0)}$ is negative, so solutions exist for which there is a mix of bosons and fermions provided the boson to fermion ratio is not too large, the bound being

$$\frac{b}{f}<\frac{\gamma_N^{(1/2)}}{|\gamma_N^{(0)}|}, \qquad N=3 \pmod 4, \qquad N<21. \tag{76}$$

For $N>21$, the $C_N^{(0)}$ are also negative so the bound is

$$\frac{b}{f}<\min\left\{\frac{C_N^{(1/2)}}{|C_N^{(0)}|}, \frac{\gamma_N^{(1/2)}}{|\gamma_N^{(0)}|}\right\}, \qquad N=3 \pmod 4, \qquad N>21. \tag{77}$$

The bound (76) is given in table 6.

TABLE 6

Upper bounds on the ratio of bosons to fermions

N	Upper bound on b/f $\dfrac{\gamma_N^{(1/2)}}{\lvert\gamma_N^{(0)}\rvert}$
3	17.47
7	23.27
11	3.523
15	0.6058
19	0.1186

What is not so satisfactory is the *value* of $\beta_N^{(j)}$ and $\gamma_N^{(j)}$. These are so small that we need an enormous number of matter fields to get reasonably small values for the gauge couplings and for $8\pi G/\rho^2$. For instance, suppose (to take a round number) that there are $f=1000$ species of fermions. The resulting values of the $O(N+1)$ and gravitational coupling parameters $g^2/8\pi^2$ and $8\pi G/8\pi^2\rho^2$ (which indicate the strength of radiative corrections) are shown in table 7. We see that even for such large numbers of matter fields, these coupling parameters are not much less than unity for any N.

We are not sure how seriously we ought to take this problem. It may be that the smallness of the coefficients $\beta_N^{(j)}$ and $\gamma_N^{(j)}$ is a special property of the sphere, which will not reappear when we turn to more realistic models with less symmetry*. Then again, maybe there really are thousands of species of matter fields.

Even if $g^2/8\pi^2$ turns out small, we are not justified in neglecting higher-order corrections if they would involve very large logarithms of energy ratios. This would be the case for instance if we used our methods to calculate the gauge coupling constants observed at ordinary energies; loops would then be associated with factors like $(g^2/8\pi^2)\ln(m_W\rho)$. The renormalization group approach teaches us that to avoid these large logarithms we should use the full $(4+N)$-dimensional theory to calculate gauge coupling constants that are renormalized at energies of order $1/\rho$, then form an effective gauge field theory by integrating out all particles of mass $\approx 1/\rho$, and only then calculate the gauge coupling constants at ordinary energies by integrating the Gell-Mann–Low equations of the effective theory down from energies

TABLE 7

Values of the $O(N+1)$ and gravitational coupling parameters for the case $f=1000$, $b=0$

N	Coupling parameters	
	$\frac{g^2}{8\pi^2}$	$\frac{G}{\pi\rho^2}$
3	0.253	0.0374
7	0.0987	0.316
11	0.426	1.72
15	1.85	8.33
19	7.96	38.3

---

* It is possible that the potential may get large contributions from one-loop graphs involving the towers of massive particles of spin 2, 1, and 0 associated with the metric itself. The numbers of these towers with spin 1 or 0 grow with N like N and N^2, respectively, so we may be able to exploit a large N instead of large b or f approximation to justify the neglect of higher graphs. Also, Duff informs us that the one-loop graphs involving massive spin-2 particles typically carry very large numerical factors, and suggests that these may dominate the one-loop potential.

of order $1/\rho$. It is unlikely that any specially simple results should be expected for fine structure constants like $\frac{1}{137}$ that appear as the end-product of this long chain of calculation.

7. Stability

We have so far only found a set of equilibrium configurations of our models, and considered their instability against growing Yang–Mills and gravitational waves. We must also ask whether the solution is stable against perturbations of the compact manifold itself.

This question can most conveniently be studied in terms of an effective potential for the whole system of matter and gravitation. In order to study stability against spacetime-independent perturbations, it is enough to consider the total action under the constraint of Poincaré invariance, which allows it to be written as

$$I_{\text{total}} = -\int \mathrm{d}^4x \, V_{\text{eff}} \,, \tag{78}$$

$$V_{\text{eff}} = \frac{1}{16\pi\bar{G}} \int \mathrm{d}^N y \sqrt{\tilde{g}(y)} [\tilde{R}(y) + \bar{\Lambda}] + V[\tilde{g}] \,. \tag{79}$$

As is well-known [13], a potential V_{eff} constructed in this way is equal to the minimum energy of any state in which the expectation value of the fields has the value indicated by the argument of the potential—in our case, $\bar{g}_{NM}(x, y)$. Therefore, provided that there are no negative-energy perturbations, a stable solution is one associated with a *minimum* of V_{eff}.

In studying perturbations which also preserve those symmetries of the manifold that keep it of the type described in appendix A (e.g. S^N, CP^N, etc.), we can regard the metric g_{nm} here as depending on a single radius parameter ρ. Eqs. (15) and (18) then let us write (79) as

$$V_{\text{eff}} = \frac{1}{16\pi G_0} \left[\frac{-N(N-1)}{\rho^2} + \bar{\Lambda} \right] + V(\rho) \,. \tag{80}$$

However it should be kept in mind that it is $\bar{G}$ that is ρ-independent, with $G_0(\rho)$ given by (18) as proportional to ρ^{-N}. Hence the derivative of V_{eff} is given by

$$\rho \frac{\mathrm{d}V_{\text{eff}}}{\mathrm{d}\rho} = \frac{N}{16\pi G_0} \left[\frac{-N(N-1)}{\rho^2} + \bar{\Lambda} \right] + \frac{2N(N-1)}{16\pi G_0 \rho^2} + \rho \frac{\mathrm{d}V}{\mathrm{d}\rho} \,.$$

We see that the field equation (23) is simply the statement that V_{eff} is stationary in ρ. (The other field equation (22) just says that $\bar{\Lambda}$ takes a value that makes V_{eff} vanish at its stationary point.) We expect that V_{eff} can also be used to judge the stability of the solution against spacetime-independent uniform dilations or contractions of the compact manifold: it is stable if and only if the stationary point of V_{eff} is a *minimum.*

Strictly speaking, this is only a guess. We really should consider the contribution to the action of terms quadratic in the time derivative of infinitesimal perturbations. We are anticipating here that these terms will as usual be negative, so that the field equations for a perturbation $\delta\rho$ will give $\delta\ddot{\rho}$ of opposite sign to $\delta\rho\, \mathrm{d}^2V/\mathrm{d}\rho^2$, and hence $\delta\rho$ will oscillate stably if $\mathrm{d}^2V/\mathrm{d}\rho^2$ is positive. This should be checked.

Returning now to $V_{\text{eff}}(\rho)$, for very small ρ we know that V_{eff} for N odd grows like C_N/ρ^4. This is the behavior when all fields have vanishing $(4+N)$-dimensional mass, and even if they do not the potential becomes independent of their masses for sufficiently small ρ. On the other hand, the curvature and cosmological constant terms in (80) go as ρ^{N-2} and ρ^N, so are negligible compared with $V(\rho)$ as $\rho\to 0$, and therefore

$$V_{\text{eff}}(\rho)\to\frac{C_N}{\rho^4}\,,\qquad \text{as } \rho\to 0\,.$$

For very large ρ we expect $V(\rho)$ to vanish as ρ^{-4} if all fields in $4+N$ dimensions are massless, or at worst approach a constant if there are $(4+N)$-dimensional masses in the theory. Here the dominant term is the cosmological term, which grows as ρ^N:

$$V_{\text{eff}}(\rho)\to\frac{\bar{\Lambda}}{16\pi G_0}\propto\rho^N\,,\qquad \text{as } \rho\to\infty\,.$$

It is clear that for C_N and $\bar{\Lambda}$ positive, the potential $V_{\text{eff}}(\rho)$ must have a minimum at some finite ρ. We found that for C_N and $\bar{\Lambda}$ positive, there is only one stationary point, and it must therefore be a minimum. Our tentative conclusion is that our solution is stable against uniform dilations or contractions. We could have found an equilibrium solution with C_N, G_0, and $\bar{\Lambda}$ all negative, but it would have been unstable.

We still have to consider stability against deformation. Page [14] has calculated the change in the C_N coefficients, for massless scalar fields, when the spherical compact manifold is subject to an aspherical but homogeneous deformation. We understand that work on the case of fermions is in progress. In fact, we would not be unhappy to learn that spheres are unstable since the high energy gauge group may very well not be an $SO(N+1)$.

Even a manifold that is stable against all deformations will become unstable if the temperature is raised above a critical value T_c, which is independent of the shape of the manifold. To see this, it is sufficient to examine the form of the free energy $F(\rho)$ for large values of the radius ρ. Suppose that the temperature has some finite value T, and consider the behavior of the total free energy when the linear scale of the manifold is much larger than $1/kT$. For definiteness, suppose that there are b species of massless scalars in 4+N dimensions. In 4 dimensions, these appear as b towers of massive particles, yielding a free energy

$$F = V_{\text{eff}} + b \sum_l D_l f(T, M_l) , \tag{81}$$

where V_{eff} is the zero temperature free energy discussed previously, and $f(T, M)$ is the thermal free energy of a scalar of mass M

$$f(T, M) = -\frac{4\pi}{3(2\pi)^3} \int_0^\infty \frac{p^4 \, \mathrm{d}p}{E_M(p)[\exp(E_M(p)/kT) - 1]} . \tag{82}$$

For a manifold whose linear scale is much larger than $1/kT$, the sum in (81) is dominated by very high eigenmodes, and can be replaced with an integral

$$F = V_{\text{eff}} + b \int_0^\infty D(M) f(T, M) \, \mathrm{d}M , \tag{83}$$

where $D(M)\,\mathrm{d}M$ is the number of eigenmodes with M_l between M and $M + \mathrm{d}M$. For instance, for spheres of radius ρ we have

$$M_l \to l/\rho ,$$
$$D_l \to 2l^{N-1}/(N-1)!$$

and so

$$D(M) = \frac{2\rho^N M^{N-1}}{(N-1)!} .$$

We can also write this in terms of the volume of the N-sphere

$$\Omega_N = \frac{2\pi^{(N+1)/2}}{\Gamma[\frac{1}{2}(N+1)]} \rho^N ,$$

so that

$$D(M) = \frac{\Omega_N M^{N-1}}{2^{N-1} \pi^{N/2} \Gamma(\frac{1}{2}N)} . \tag{84}$$

The point of writing the formula this way is that in this form it is valid for manifolds of arbitrary shape, not just spheres [15]. It is now elementary to substitute (82) and (84) in (83), and do the integrals over M and p/M (in that order). The result is

$$F = V_{\text{eff}} - \frac{b(N+3)!\zeta(N+4)\Omega_N (kT)^{N+4}}{2^{N+3}\pi^{(N+3)/2}\Gamma[\frac{1}{2}(N+5)]} . \tag{85}$$

This formula applies (for large manifolds) even to the free energy generated by the massive excitations of the metric itself, with

$$b_{\text{grav}} = \tfrac{1}{2}(N+4)(N+5) - 2(N+4) = \tfrac{1}{2}(N+1)(N+4) .$$

The negative $(kT)^5\rho$ term contributed by the excitations of the metric has already been found in the original Kaluza–Klein model (where $b_{\rm grav}=5$) by Rubin and Roth [16]. It is easy to see that (85) holds as long as the total dimensionality is $N+4$, whatever the dimensionalities of the compact manifold and Minkowski subspace.

Comparing (85) with (80) and recalling that $G_0=\bar{G}/\Omega_{\rm N}$, we see that the cosmological constant term dominates over the thermal term for large manifolds only for T less than a critical temperature T_c, given by

$$\frac{\bar{A}}{16\pi\bar{G}}=\frac{b(N+3)!\,\zeta(N+4)(kT_c)^{N+4}}{2^{N+3}\pi^{(N+3)/2}\Gamma[\frac{1}{2}(N+5)]}. \tag{86}$$

This gives kT_c roughly of order $1/\sqrt{bG}$. For T greater than T_c, the free energy (85) becomes increasingly negative for $\rho\to\infty$, and we no longer have even an equilibrium solution.

This suggests that there is a dramatic phase transition at $T=T_c$, in which the compactified dimensions explode outward. To see what happens then, it will clearly be necessary to give up our assumption of Poincaré invariance, and include at least a time dependence in our ansatz for the $(4+N)$-dimensional metric. This phase transition presumably will come into play in the interior of black holes (though perhaps not in any finite time as seen by outside observers) and perhaps also in the last stages of black hole evaporation and cosmic contraction. One wonders also whether our universe started with equal circumferences in all $3+N$ spatial directions, and became tightly contracted in N of those dimensions only when the temperature fell below the critical temperature T_c. Work on these problems is in progress.

We have been greatly helped throughout the course of this work by frequent enlightening conversations with M. Duff, D. Page, and M. Rubin. Thanks for valuable discussions are also due to B. DeWitt, H. Georgi, P. Ginsparg, C. Pope, D. Toms and E. Witten. We wish to thank M. Awada and D. Toms for bringing to our attention errors with regard to the calculation of the $D_N^{(1/2)}$ and $E_N^{(1/2)}$ coefficients that were present in an earlier version of this article and for communicating their work prior to publication.

Appendix A

ONE-PARAMETER MANIFOLDS

In this appendix we identify a class of Riemmanian manifolds whose symmetries dictate that:

(i) the metric has only a single free parameter ρ, which appears (in a suitable coordinate system) as a multiplicative factor ρ^2;

(ii) all symmetric second-rank tensors that share the symmetries of the metric are proportional to the metric, with constant coefficients.

By a tensor S_{nm} sharing the symmetries of the metric, we mean that for any Killing vector $\xi^n(y)$ that satisfies the usual condition

$$\xi_{n;m} + \xi_{m;n} = 0\,, \tag{A.1}$$

the Lie derivative of S with respect to ξ vanishes [17]:

$$\xi^p S_{nm;p} + S_{pm}\xi^p{}_{;n} + S_{np}\xi^p{}_{;m} = 0\,. \tag{A.2}$$

(Tildes are dropped in this appendix.) We will see that (as is pretty obvious) the matter energy-momentum tensor T^{nm} and the Ricci tensor R^{nm} share the symmetries of the metric in this sense, so property (ii) allows us to replace the $\frac{1}{2}N(N+1)$ differential equations (12) with a single algebraic equation.

We understand that the conditions for a manifold to satisfy properties (i) and (ii) are well known to mathematicians [18]. However, we have not been able to find a reference for this, and we felt that a discussion here might be of some use to physicists working on this sort of theory. (See added note below.)

The manifolds to be considered here are homogeneous, with a group of isometries G. By "homogeneous" we mean that any point is carried into any other point by some element of G. For any specific point y_0, there may be some subgroup H of G that leaves y_0 fixed. (Homogeneity ensures that any other point is left fixed by a subgroup isomorphic to H.) Any other point y can be obtained from y_0 by some transformation g in G, so points in the manifold can be labelled by the elements of G, two elements g_1, g_2 of G corresponding to the same point of the manifold if $g_1^{-1}g_2$ is in H. That is, elements of the manifold are in one-to-one correspondence with the cosets that make up the coset space G/H. It is often convenient to split the generators of G into generators T_i of H and a linearly independent set X_n outside the algebra of H, and define the coordinates y^n of the manifold (at least locally) by specifying that y_0 is carried into y by the G-element $\exp(iy^n X_n)$.

Let us first apply the invariance of the metric and other tensors S_{nm} under the isotropy subgroup H. An infinitesimal element $h = 1 + i\varepsilon^i T_i$ of H induces a coordinate transformation

$$y^n \to y'^n = y^n + \varepsilon^i \xi^n{}_i(y)\,, \tag{A.3}$$

with ξ^n_i a set of Killing vectors corresponding to the generators T_i of H. Since H leaves y_0 fixed, these Killing vectors vanish as y_0, so (A.2) becomes at y_0:

$$S_{pm}(y_0)(t_i)^p{}_n + S_{np}(y_0)(t_i)^p{}_m = 0\,, \tag{A.4}$$

where

$$(t_i)^p{}_n \equiv (\partial_n \xi^p_i)_{y=y_0}\,. \tag{A.5}$$

The matrices t_i are just the generators of H in the representation D_X furnished by the coordinates, i.e. by the generators X_n of G that are not in the subgroup H. If the symmetric direct product $(\mathrm{D}_X \otimes D_X)_s$ contains the identity representation just once, then (A.4) gives

$$S_{nm}(y^0) = \mathcal{S}\Delta_{nm}\,, \tag{A.6}$$

where Δ_{nm} is the Clebsch–Gordan coefficient for forming a singlet from $(\mathrm{D}_X \otimes \mathrm{D}_X)_\mathrm{s}$, and $\mathcal{S}$ is an unknown constant.

Now we apply the invariance of these tensors under finite transformations in G that do not leave y_0 fixed. In general, the integrated form of (A.2) is

$$S_{nm}(y') = S'_{nm}(y') \equiv \frac{\partial y^p}{\partial y'^n}\frac{\partial v^q}{\partial y'^m}S_{pq}(y)\,, \tag{A.7}$$

where $y \to Y'(y, g)$ is the transformation induced by some element g of G. Setting $y = y_0$ and taking g as $\exp(iy^nX_n)$ (with a new y^n), we find

$$S_{nm}(y) = \mathcal{S}(B^{-1}(y))^p{}_n(B^{-1}(y))^q{}_m\Delta_{pq} \tag{A.8}$$

where $B(y)$ is the "boost" matrix

$$B^n{}_m(y) = \frac{\partial}{\partial y_0^m}y'^n(y_0, \exp(iy^nX_n))\,. \tag{A.9}$$

The important thing about (A.8) is that it shows that all symmetric tensors that satisfy the invariance conditions (A.2) are equal, up to multiplicative constant factors. This is in particular true of the metric tensor (for which (A.1) and (A.2) are the same) so the metric tensor is uniquely specified by its symmetries and our choice of coordinates up to a multiplicative constant, and all other tensors satisfying (A.2) (including the Ricci tensor) are proportional to the metric.

The special assumption that led to this result was that the manifold is a coset space G/H for which the generators of G not in H form a representation D_X of H such that the symmetric direct product $(\mathrm{D}_X \times \mathrm{D}_X)_\mathrm{s}$ contains the identity just once. The real representation D_X will have this property if and only if it is either irreducible, or the sum of an irreducible complex representation and its complex conjugate. Let us see how this works in some familiar cases.

Spheres: The sphere S^N can be described as the coset space $\mathrm{SO}(N+1)/\mathrm{SO}(N)$. The generators X_n of $\mathrm{SO}(N+1)$ which are not in $\mathrm{SO}(N)$ for a single N-vector representation of $\mathrm{SO}(N)$. Since this is irreducible, properties (i) and (ii) hold for spheres, as is of course well known.

CP^N: The space CP^N can be described as the coset space $\mathrm{SU}(N+1)/\mathrm{SU}(N)\times \mathrm{U}(1)$. The generators of $\mathrm{SU}(N+1)$ which are not in $\mathrm{SU}(N)\times U(1)$ form two representations $\boldsymbol{N}$ and $\bar{\boldsymbol{N}}$ of $\mathrm{SU}(N)$, with opposite values for the $U(1)$ quantum number. Since $\boldsymbol{N}$ is an irreducible complex representation of $\mathrm{SU}(N)\times \mathrm{U}(1)$, properties (i) and (ii) hold also for CP^N.

Group manifolds: We can endow the manifold of a compact Lie group G with a symmetry under the group $\mathrm{G}\times\mathrm{G}$, an element g of G being carried by an element

(a, b) of G×G into agb^{-1}. The isotropy group H that leaves the particular element $g=g_0$ of G invariant is $\mathrm{G}_{\mathrm{diag}}$, consisting of elements (a, b) of G×G with $b=g_0{}^{-1}ag_0$, so the manifold of G can be described as the coset space $\mathrm{G}\times\mathrm{G}/\mathrm{G}_{\mathrm{diag}}$. If A_n and B_n are the generators of the left and right G factors of G×G, then the generators of $\mathrm{H}=\mathrm{G}_{\mathrm{diag}}$ are $A_n+g_0^{-1}B_ng_0$, and the generators X_n of G×G that are not in $\mathrm{G}_{\mathrm{diag}}$ can be taken as just A_n. The A_n transform according to the adjoint representation of $\mathrm{G}_{\mathrm{diag}}$, which for a simple group G is real and irreducible, so properties (i) and (ii) hold for simple group manifolds. On the other hand, if the group G is the direct product of $K>1$ simple subgroups, then $(\mathrm{D}_X\otimes\mathrm{D}_X)_{\mathrm{s}}$ contains the identity K times, and properties (i) and (ii) do not hold.

The squashed 7-sphere: The squashed 7-sphere can be described as the coset space SO(5)×SO(3)/SO(3)×SO(3), with the generators of SO(5)×SO(3) that are not in SO(3)×SO(3) transforming as the SO(3)×SO(3) representation (1,3)+(2,2) [19]. There are here two ways of forming a singlet, from (1,3)×(1,3) or (2,2)×(2,2), so (as is well known) properties (i) and (ii) do not hold for the squashed 7-sphere.

There is a generalization of properties (i) and (ii), which applies to all homogeneous spaces G/H. If the elements of G that are not in H form a representation D_X for which the symmetric direct product $(\mathrm{D}_X\otimes\mathrm{D}_X)_{\mathrm{s}}$ contains the identity K times, then (A.8) is replaced with

$$S_{nm}(y)=\sum_{j=1}^{K}\mathcal{S}_j(B^{-1}(y))^p{}_n(B^{-1}(y))^q{}_m\Delta^{(j)}_{pq}\,,$$

with $\Delta^{(j)}$ a set of Clebsch–Gordan coefficients that depend only on the manifold and our coordinate system, and $\mathcal{S}_j$ a set of constant coefficients. The metric and all other tensors satisfying (A.2) are thus linear combinations of just K standard tensors. This is in particular true for the manifolds of semi-simple groups, for which K is the number of simple factors, and for the squashed 7-sphere, for which $K=2$.

Finally we note that one class of tensors that automatically share the symmetries of the metric are those formed by variational derivatives of scalar functionals $F[g]$ with respect to the metric:

$$\tfrac{1}{2}\sqrt{g(y)}S^{nm}(y)\equiv\frac{\delta F[g]}{\delta g_{nm}(y)}\,. \tag{A.10}$$

This can be proved formally by noting that the general covariance of $F[g]$ implies that S^{nm} is conserved for all metrics, and therefore

$$\int \xi_n S^{nm}{}_{;m}\sqrt{g}\,\mathrm{d}^Ny=0\,,$$

for *all* g_{nm}, whether or not ξ_n is a Killing vector of g_{nm}. A small variation in g_{nm}

(with ξ_n fixed) thus satisfies

$$\int \xi_n (\delta S^{nm})_{;m} \sqrt{g}\, \mathrm{d}^N y + \int \xi_n S^{pm} \sqrt{g}\, \delta \Gamma^n_{pm}\, \mathrm{d}^N y = 0\,.$$

Now fix g_{nm} to be a metric with Killing vector ξ_n. Integrating the first term here by parts and using (A.1) and the symmetry of S^{nm}, we see that this term vanishes. Also, the variation in the affine connection can be expressed as

$$\delta \Gamma^n_{pm} = \tfrac{1}{2} g^{nq} [(\delta g_{qp})_{;m} + (\delta g_{qm})_{;p} - (\delta g_{pm})_{;q}]\,.$$

Using this in the second term above and integrating by parts then yields

$$0 = \int [\xi^m S^{pq}{}_{;m} - \xi^q{}_{;m} S^{mp} - \xi^p{}_{;m} S^{mq}] \sqrt{g}\, \delta g_{qp}\, \mathrm{d}^N y\,.$$

Since this holds for all δg_{qp}, the coefficient of δg_{qp} must vanish which (using (A.1) again) is equivalent to (A.2).

The Einstein tensor $R^{nm} - \frac{1}{2} g^{nm} R$ and the matter-fluctuation energy-momentum tensor T^{nm} are of the type (A.10), with $F[g]$ respectively the action $\int \mathrm{d}^N y \sqrt{g} R(y)$ and the potential V. Thus, without any calculation, we know that S^N, CP^N, etc. are Einstein spaces, and that the field equations reduce for each such manifold to a single algebraic equation for the radius ρ.

Added Note. The results regarding one-parameter manifolds discussed in appendix A were earlier derived by Zee [20], and Boulware and Brown [21]. It may be useful to note further that a homogeneous space G/H can satisfy the irreducibility condition for a compact manifold (i.e. that the generators of G not in the algebra of H form a representation of H that is irreducible into real representations) only if H is a maximal subgroup of G (i.e., there is no subgroup of G that contains H, other than G or H). Also, the irreducibility condition *will* be satisfied if H is maximal and G/H is a symmetric space (in the sense that G has an automorphism that leaves the generators of H unchanged and reverses the sign of the other generators of G); see Helgason [22], Peskin [23].

Appendix B

THE EFFECTIVE POTENTIAL FOR $\mathrm{M}^4 \times \mathrm{S}^N$

In this appendix we shall compute the effective potential for the scalar and Dirac fields for a background geometry which is the direct product of a four-dimensional Minkowski spacetime with the N-sphere.

Minimally coupled scalar fields

We shall take the action for a minimally coupled scalar field to be

$$S[\phi] = -\tfrac{1}{2} \mu^{n-4} \int \mathrm{d}^n x \int \mathrm{d}^N y\, g^{1/2} (\phi_{;A} \phi_{;}^{A} + m^2 \phi^2)\,, \tag{B.1}$$

where n, in the sense of dimensional regularization, denotes the dimension of spacetime, upper case indices run over the values $0, 1, \dots, n+N-1$, and μ is an arbitrary constant with the dimension of mass.

The effective potential V is defined through the relation

$$\exp\left(-i\mu^{n-4}\int \mathrm{d}^n x\, V\right) = \int \mathscr{D}\phi\, \mathrm{e}^{iS[\phi]}\,. \tag{B.2}$$

It follows that

$$\exp\left(-i\mu^{n-4}\int \mathrm{d}^n x\, V\right) = \det{}^{-1/2}(-\Box + m^2)\,, \tag{B.3}$$

and hence that

$$V = -\tfrac{1}{2} i\mu^{-(n-4)} \int \frac{\mathrm{d}^n k}{(2\pi)^n} \sum_{l=0}^{\infty} D_l \log(k^2 + \Lambda_l^2 \rho^{-2} + m^2)\,, \tag{B.4}$$

where $-\Lambda_l^2$ and D_l denote the eigenvalues and the degeneracies of the laplacian on the unit N-sphere. These are known to be [9]

$$\Lambda_l^2 = l(l+N-1)\,, \qquad D_l = (2l+N-1)\frac{\Gamma(l+N-1)}{\Gamma(N)\,l!}\,, \tag{B.5}$$

where we express the D_l in a form that anticipates that it will be convenient in the following to regard N as a continuous variable. A short calculation reveals that

$$\int \frac{\mathrm{d}^n k}{(2\pi)^n} \log(k^2 + m_l^2) = -\frac{i\Gamma(-\frac{1}{2}n) m_l^n}{(4\pi)^{n/2}}\,. \tag{B.6}$$

On using this result in (4) we find

$$V = -\frac{\mu^{-(n-4)}\Gamma(-\frac{1}{2}n)}{(4\pi)^{n/2}\rho^n} \sum_{l=0}^{\infty} \frac{\Gamma(l+N-1)}{l!\,\Gamma(N)} (l+\tfrac{1}{2}(N-1))(l(l+N-1)+m^2\rho^2)^{n/2}\,, \tag{B.7}$$

subject to the restriction $\mathrm{Re}\,(N+n)<0$ which is required to secure the convergence of the sum.

We seek now to simplify the sum and to analytically continue V to the required values of n and N. To this end we note that the summand in (B.7) is a binomial coefficient multiplied by the troublesome term

$$\begin{aligned}(l+\tfrac{1}{2}(N-1))&[l(l+N-1)+m^2\rho^2]^{n/2}\\ &= (l+\tfrac{1}{2}(N-1))[(l+\tfrac{1}{2}(N-1))^2 - ((\tfrac{1}{2}(N-1))^2 - m^2\rho^2)]^{n/2}\,.\end{aligned} \tag{B.8}$$

Our strategy will be to replace this quantity by its representation as a Laplace integral; this will enable us to perform the sum. The following identity proves useful in this regard

$$z(z^2-b^2)^{n/2}=\frac{\pi^{1/2}2^{n/2+1/2}b^{n/2+3/2}}{\Gamma(-\frac{1}{2}n)}\int_0^\infty \mathrm{d}t\,\mathrm{e}^{-zt}t^{-n/2-1/2}I_{-n/2-3/2}(bt)\,,\qquad \mathrm{Re}\,n<-1 \tag{B.9}$$

with I a modified Bessel function [9]. On using (B.9) in (B.7) and taking note of the fact that

$$\begin{aligned}\sum_{l=0}^{\infty}\frac{\Gamma(l+N-1)}{l!}\,\mathrm{e}^{-(l+\frac{1}{2}(N-1))t}&=\Gamma(N-1)\,\mathrm{e}^{-(N-1)t/2}(1-\mathrm{e}^{-t})^{-(N-1)}\\&=\Gamma(N-1)(2\,\mathrm{sh}\,\tfrac{1}{2}t)^{-(N-1)}\,,\end{aligned} \tag{B.10}$$

we find

$$V=-\frac{\mu^{-(n-4)}}{(2\pi)^{n/2-1/2}(N-1)\rho^n}\int_0^\infty \mathrm{d}t\,t^{-N-n-1}f(t^2)\,,\qquad \mathrm{Re}\,(N+n)<0\,, \tag{B.11}$$

where we have defined

$$\begin{aligned}f(t^2)&=\left(\frac{t}{2\mathrm{sh}\frac{1}{2}t}\right)^{N-1}(bt)^{n/2+3/2}I_{-n/2-3/2}\,(bt)\,,\\ b^2&=[\tfrac{1}{2}(N-1)]^2-\ m^2\rho^2\,,\end{aligned} \tag{B.12}$$

which we observe to be a function analytic in t^2 about $t^2=0$.

It remains now to continue the integral in (B.11) to the physical values of N and n. One way to proceed would be to integrate by parts repeatedly. However a simpler method presents itself. Consider the integral

$$\int_{\mathrm{C}} \mathrm{d}t\,t^{-N-n-1}f(t^2)\,,$$

with C the contour of fig. 1. By displacing the contour to C′, which runs just above the branch cut associated with t^{-N-n-1}, we see that

$$\int_{\mathrm{C}} \mathrm{d}t\,t^{-N-n-1}f(t^2)=(1-\mathrm{e}^{\ i\pi(N+n)})\int_0^\infty \mathrm{d}t\,t^{-N-n-1}f(t^2)\,, \tag{B.13}$$

i.e.

$$\int_0^\infty \mathrm{d}t\,t^{-N-n-1}f(t^2)=\frac{1}{(1-\mathrm{e}^{-i\pi(N+n)})}\int_{\mathrm{C}} \mathrm{d}t\,t^{-N-n-1}f(t^2)\,. \tag{B.14}$$

We observe that since the contour C does not run through the point $t=0$ the integral over C exists and is analytic for all values of $N+n$, the poles of the left-hand side

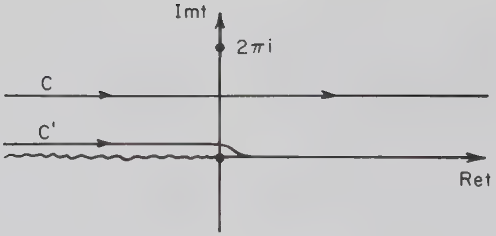


FIG. 1.

being contained in the factor

$$\frac{1}{(1-e^{-i\pi(N+n)})}$$

The right-hand side of (B.14) represents, therefore, the analytic continuation of the left-hand side to the entire $N+n$ plane with the exception of the points $N+n=0, 2, 4, \ldots$. In order to avoid the poles for $N+n$ even we shall set $n=4$ and take N to be an odd integer. It follows from (B.12) and (B.14) that

$$V=-\frac{1}{2(2\pi)^{3/2}(N-1)\rho^4}\int_C \frac{dt}{t^6}\frac{(bt)^{7/2}I_{-7/2}(bt)}{(2\,\mathrm{sh}\tfrac{1}{2}t)^{N-1}}, \qquad n=4, \qquad N \text{ odd}. \tag{B.15}$$

We specialize now to the case $m=0$ for which

$$b=\tfrac{1}{2}(N-1)\equiv\nu, \qquad m=0. \tag{B.16}$$

Note also that since the Bessel function is of half odd-integer order it is related to hyperbolic functions. Specifically

$$z^{7/2}I_{-7/2}(z)=\sqrt{2/\pi}\{(15z+z^3)\,\mathrm{sh}\,z-(15+6z^2)\,\mathrm{ch}\,z\}, \tag{B.17}$$

and hence

$$V=-\frac{1}{(2\pi)^2 2\nu\rho^4}\int_C \frac{dt}{t^6}\frac{1}{(2\,\mathrm{sh}\tfrac{1}{2}t)^{2\nu}}$$
$$\times\{(15\nu t+(\nu t)^3)\,\mathrm{sh}\,(\nu t)-(15+6(\nu t)^2)\,\mathrm{ch}\,(\nu t)\}. \tag{B.18}$$

The integrand has poles corresponding to the zeros of $\mathrm{sh}\,\frac{1}{2}t$ i.e. when $t=\pm 2p\pi i$, $p=1, 2, \ldots$ and may, in principle, be evaluated by residues which leads to sums of Riemann ζ functions. Thus, for example, we find for $\nu=1$ (i.e. $N=3$)

$$\rho^4 V=\frac{1}{2}\left(\frac{\zeta(3)}{(2\pi)^4}-\frac{39\zeta(5)}{(2\pi)^6}+\frac{90\zeta(7)}{(2\pi)^8}\right)=7.56870\times 10^{-5},$$
$$n=4, \qquad N=3, \qquad m=0. \tag{B.19}$$

However the computations become arduous if ν is large. We prefer, therefore, to cast (B.18) in a form that is amenable to numerical integration. We take C to be the contour $(-\infty+\pi i, \infty+\pi i)$. On setting

$$t=\pi(i+\tan\theta)\,, \tag{B.20}$$

and availing ourselves of the elementary identity

$$\int_{-\pi/2}^{\pi/2} d\theta\, h(\theta)=\int_0^{\pi/2} d\theta\,(h(\theta)+h(-\theta))\,, \tag{B.21}$$

we obtain our final expression for the effective potential

$$\begin{aligned} V=\frac{1}{4\pi\nu\rho^4}\int_0^{\pi/2}\frac{d\theta}{[2\,\mathrm{ch}\,(\frac{1}{2}\pi\tan\theta)]^{2\nu}}&\left\{\left(\frac{\nu^3}{\pi^3}\cos\theta\sin 3\theta-\frac{15\nu}{\pi^5}\cos^3\theta\sin 5\theta\right)\right.\\ &\left.\times\,\mathrm{sh}\,(\nu\pi\tan\theta)+\left(\frac{6\nu^2}{\pi^4}\cos^2\theta\cos 4\theta-\frac{15}{\pi^6}\cos^4\theta\cos 6\theta\right)\mathrm{ch}\,(\nu\pi\tan\theta)\right\},\\ n=4\,,\qquad N\ \text{odd}\,,\qquad m=0\,. \end{aligned} \tag{B.22}$$

The case $\nu=0$ (i.e. $N=1$) is best dealt with by returning to (B18) and taking the limit $\nu\to 0$. Since

$$\int_C \frac{dt}{t^6}=0\,, \tag{B.23}$$

we find in this case that

$$V=-\frac{15}{4\pi^2\rho^4}\int_C \frac{dt}{t^6}\log(\mathrm{sh}\,(\tfrac{1}{2}t))\quad . \tag{B.24}$$

Integration by parts and evaluation by residues immediately yields

$$V=-\frac{3\zeta(5)}{(2\pi)^6\rho^4}=-\frac{5.05576\times 10^{-5}}{\rho^4}\,,\qquad n=4\,,\qquad N=1\,,\qquad m=0 \tag{B.25}$$

in agreement with the result of ref. [5].

Conformally coupled scalar fields

The case of conformal coupling is dealt with by making the replacement

$$m^2\to\frac{N(N+2)(N-1)}{4(N+3)\rho^2}\,, \tag{B.26}$$

or equivalently, the replacement

$$b\to i\gamma\,,\qquad \gamma^2\equiv\frac{3}{4}\left(\frac{N-1}{N+3}\right), \tag{B.27}$$

in (15). This leads rapidly to the expression

$$C_N^{\text{conformal}}=\frac{(-1)^{\nu+1}}{8\pi\nu}\int_0^{\pi/2}\frac{\mathrm{d}\theta}{[2\,\mathrm{ch}\,(\frac{1}{2}\pi\tan\theta)]^{2\nu}}$$

$$\times\left\{\frac{15\gamma}{\pi^5}\cos^3\theta(\mathrm{e}^{-\pi\gamma}\cos(5\theta+\pi\gamma\tan\theta)-\mathrm{e}^{\pi\gamma}\cos(5\theta-\pi\gamma\tan\theta))\right.$$

$$+\frac{\gamma^3}{\pi^3}\cos\theta(\mathrm{e}^{-\pi\gamma}\cos(3\theta+\pi\gamma\tan\theta)-\mathrm{e}^{\pi\gamma}\cos(3\theta-\pi\gamma\tan\theta))$$

$$+\frac{15}{\pi^6}\cos^4\theta(\mathrm{e}^{-\pi\gamma}\cos(6\theta+\pi\gamma\tan\theta)+\mathrm{e}^{\pi\gamma}\cos(6\theta-\pi\gamma\tan\theta))$$

$$\left.+\frac{6\gamma^2}{\pi^4}\cos^2\theta(\mathrm{e}^{-\pi\gamma}\cos(4\theta+\pi\gamma\tan\theta)+\mathrm{e}^{\pi\gamma}\cos(4\theta-\pi\gamma\tan\theta))\right\},$$

$$n=4,\ N\text{ odd},\ m=0,\ \text{conformal coupling}. \qquad \text{(B.28)}$$

Numerical integration yields the results of table 3. A check on the numerical work is provided by evaluating the integral by residues for the case $\nu=1$ (i.e. $N=3$) this yields the expression

$$C_3=-\frac{1}{107520\pi}+\frac{1}{512\pi}\sum_{p=1}^{\infty}\frac{\mathrm{e}^{-p\pi}}{(p\pi)^3}\left(1+\frac{9}{p\pi}+\frac{39}{(p\pi)^2}+\frac{90}{(p\pi)^3}+\frac{90}{(p\pi)^4}\right)$$

$$=7.145890079\times10^{-6}.$$

$$n=4,\ N=3,\ m=0,\qquad \text{conformal coupling}. \qquad \text{(B.29)}$$

These values of C_n can be used to calculate the potential V for the case of conformal coupling, but in this case there are other changes in the gravitational field equations, which we shall not consider in this paper.

Fermions

The calculation of the effective potential for fermions is a problem of a slightly more recondite nature than the corresponding scalar problem since, in order to find the expression analogous to (B.3), we must first compute the eigenvalues and the degeneracies of the Dirac laplacian on S^N. We shall therefore calculate V via the Green function method and discover the eigenvalues and their degeneracies as we proceed.

Our conventions regarding the Dirac matrices, $\hat{\gamma}^A$, for $\mathrm{M}^4\times\mathrm{S}^N$ are as follows

$$\{\hat{\gamma}^A,\hat{\gamma}^B\}=2g^{AB},$$

$$\hat{\gamma}^\mu=\gamma^\mu\otimes\mathbb{1},\qquad \hat{\gamma}^p=\gamma^5\otimes\Gamma^p,\qquad (\gamma^5)^2=+1, \qquad \text{(B.30)}$$

where the $\gamma^\mu(\mu=0,1,2,3)$ are Dirac matrices appropriate to M^4 while the $\Gamma^p(p=4,5,\ldots,4+N-1)$ are those appropriate to S^N.

We take the Dirac action to be

$$S[\psi]=i\mu^{n-4}\int \mathrm{d}^n x\int \mathrm{d}^N y\, g^{1/2}\bar{\psi}(\hat{\gamma}^A\psi_{;A}+m)\psi\,, \tag{B.31}$$

and define the Green function and the effective potential by

$$G(x,y;x',y')=i\mu^{n-4}\frac{\int \mathscr{D}\psi\mathscr{D}\bar{\psi}\psi(x,y)\bar{\psi}(x',y')\,\mathrm{e}^{iS[\psi]}}{\int \mathscr{D}\psi\mathscr{D}\bar{\psi}\,\mathrm{e}^{iS[\psi]}}\,, \tag{B.32}$$

$$\exp\left(-i\mu^{n-4}\int \mathrm{d}^n x\, V\right)=\int \mathscr{D}\psi\mathscr{D}\bar{\psi}\,\mathrm{e}^{iS[\psi]}\,. \tag{B.33}$$

By differentiating V with respect to m we see that the effective potential is related to the coincident limit of the Green function by the equation

$$\frac{\partial V}{\partial m}=\mu^{-(n-4)}\Omega_N\,\mathrm{Tr}\,G(x,y;x,y)\,, \tag{B.34}$$

where Tr indicates a trace over spinor indices.

It is convenient to resolve the x, x' dependence of G by means of Fourier integral

$$G(x,y;x',y')=\int\frac{\mathrm{d}^n k}{(2\pi)^n}\,\mathrm{e}^{ik(x-x')}G(y,y'|k)\,, \tag{B.35}$$

and hence we have for the coincident limit

$$G(x,y;x,y)=\int\frac{\mathrm{d}^n k}{(2\pi)^n}\,G(y,y|k)\,. \tag{B.36}$$

We wish now to relate the Green function on $\mathrm{M}^n\times\mathrm{S}^N$ to that on S^N. $G(x,y;x',y')$ satisfies the equation

$$-i(\hat{\gamma}^A\mathscr{D}_A+m)G(x,y;x',y')=-g^{-1/2}\delta(x,x')\delta(y,y')\,, \tag{B.37}$$

and hence $G(y,y'|k)$ satisfies

$$(-i(m+i\not{k})\otimes\mathbb{1}+\gamma^5\otimes(-i\Gamma^p\mathscr{D}_p))G(y,y'|k)=-g^{-1/2}\delta(y,y')\,. \tag{B.38}$$

To solve this equation we shall decompose $G(y,y'|k)$ according to the action of γ^5. In order to do this we introduce the matrices

$$U_\pm=\frac{1}{2\sqrt{k^2+m^2}}(\gamma^5\sqrt{k^2+m^2}\pm(m-i\not{k}))\,. \tag{B.39}$$

These have the properties that

$$(m+i\not{k})U_\pm=\pm\sqrt{k^2+m^2}\,\gamma^5U_\pm\,, \tag{B.40}$$

$$\gamma^5(U_++U_-)=\mathbb{1}\,. \tag{B.41}$$

In view of the above we set

$$G(y, y'|k) = U_+ \otimes G_+(y, y'|k) + U_- \otimes G_-(y, y'|k) . \tag{B.42}$$

On substituting this ansatz into (B.38) we find that

$$\begin{aligned}(\gamma^5 U_+) \otimes [-i(\Gamma^p \mathscr{D}_p + \sqrt{k^2+m^2}) G_+] \\ + (\gamma^5 U_-) \otimes [-i(\Gamma^p \mathscr{D}_p - \sqrt{k^2+m^2}) G_-] = -g^{-1/2} \delta(y, y') .\end{aligned} \tag{B.43}$$

Therefore the ansatz (B.42) solves (B.38) provided the $G_\pm$ satisfy the equations

$$-i(\Gamma^p \mathscr{D}_p \pm \sqrt{k^2+m^2}) G_\pm(y, y'|k) = -g^{-1/2} \delta(y, y') , \tag{B.44}$$

i.e. the $G_\pm$ are the spinor Green functions on S^N.

Referring back to (B.34) and noting that

$$\operatorname{Tr} U_\pm = \frac{m \operatorname{tr} 1}{2\sqrt{k^2+m^2}} , \tag{B.45}$$

where tr 1 indicates the dimensionality of the γ^μ, we see that

$$\frac{\partial V}{\partial m} = \tfrac{1}{2} \mu^{-(n-4)} \Omega_N (\operatorname{tr} 1) \int \frac{\mathrm{d}^n k}{(2\pi)^n} \frac{m}{\sqrt{k^2+m^2}} (\operatorname{tr} G_+(y, y|k) - \operatorname{tr} G_-(y, y|k)) . \tag{B.46}$$

The coincident limits of the $G_\pm$ are known from the results of ref. [24] to be given by

$$\operatorname{tr} G_\pm = \pm i (\operatorname{tr} \mathbb{1}) \frac{1}{(4\pi)^{N/2} \rho^{N-1}} \beta \frac{\Gamma(\frac{1}{2}N + i\beta) \Gamma(\frac{1}{2}N - i\beta)}{\Gamma(1+i\beta) \Gamma(1-i\beta)} \Gamma(1 - \tfrac{1}{2}N) , \qquad \operatorname{Re} N < 2 \tag{B.47}$$

where $\operatorname{tr} \mathbb{1}$ indicates the dimensionality of the Γ^p and

$$\beta = \rho \sqrt{k^2+m^2} . \tag{B.48}$$

If we substitute (B.47) into (B.46) and recall the duplication formula

$$\Gamma(z) \Gamma(z + \tfrac{1}{2}) = \pi^{1/2} 2^{-2z+1} \Gamma(2z) , \tag{B.49}$$

we find

$$\begin{aligned} V = {} & \frac{\mu^{-(n-4)}}{(4\pi)^{n/2}} \operatorname{Tr} (1 \otimes \mathbb{1}) \frac{\sin \frac{1}{2} n\pi}{\sin \frac{1}{2} N\pi} \frac{\Gamma(1 - \frac{1}{2} n)}{\Gamma(N)} \\ & \times \int_0^\infty \mathrm{d}k\, k^{n-1} \int_{(k^2+m^2)\rho^2}^\infty \mathrm{d}\beta^2 \frac{\Gamma(\frac{1}{2}N + i\beta) \Gamma(\frac{1}{2}N - i\beta)}{\Gamma(1+i\beta) \Gamma(1-i\beta)} \end{aligned} \tag{B.50}$$

$$\begin{aligned} = {} & -\frac{\mu^{-(n-4)}}{(4\pi)^{n/2}} \operatorname{Tr} (1 \otimes \mathbb{1}) \frac{\sin \frac{1}{2} n\pi}{\sin \frac{1}{2} N\pi} \frac{\Gamma(-\frac{1}{2} n)}{\Gamma(N)} \rho^2 \\ & \times \int_0^\infty \mathrm{d}k\, k^{n+1} \frac{\Gamma(\frac{1}{2}N + i\beta) \Gamma(\frac{1}{2}N - i\beta)}{\Gamma(1+i\beta) \Gamma(1(-i\beta)} , \qquad \operatorname{Re} N + n < 0 , \qquad \operatorname{Re} n > -2 \end{aligned} \qquad \text{(B.51}$$

$$
\begin{aligned}
&= \tfrac{1}{2} i \frac{\mu^{-(n-4)}}{(4\pi)^{n/2}} \operatorname{Tr}(1 \otimes \mathbb{1}) \frac{\mathrm{e}^{-in\pi/2}}{\sin \frac{1}{2} N\pi} \frac{\Gamma(-\frac{1}{2}n)}{\Gamma(N)} \rho^2 \\
&\quad \times \int_C \mathrm{d}k\, k^{n+1} \frac{\Gamma(\frac{1}{2}N + i\beta)\Gamma(\frac{1}{2}N - i\beta)}{\Gamma(1+i\beta)\Gamma(1-i\beta)}, \qquad \operatorname{Re} N + n < 0 .
\end{aligned}
\tag{B.52}
$$

The second equality following by an integration by parts and the third by replacing the contour of integration by one similar to that of fig. 1. The simplest way to proceed seems to be to note that the integrand in (B.52) has poles when

$$
i\beta = \pm(\tfrac{1}{2}N + l), \qquad l = 0, 1, 2, \ldots, \tag{B.53}
$$

and to evaluate the integral by residues. This yields an expression analogous to (B.7)

$$
V = \frac{\mu^{-(n-4)}}{(4\pi)^{n/2} \rho^n} \operatorname{Tr}(1 \otimes \mathbb{1}) \Gamma(-\tfrac{1}{2}n) \sum_{l=0}^{\infty} \frac{\Gamma(l+N)}{l!\,\Gamma(N)} (m^2\rho^2 + (l + \tfrac{1}{2}N)^2)^{n/2}, \qquad \operatorname{Re}(N+n) < 0 . \tag{B.54}
$$

Comparing this sum with (B.7) we see that the eigenvalues of the spinor laplacian on S^N are, for $l = 0, 1, 2, \ldots,$

$$
-\lambda_l^2 = -(l + \tfrac{1}{2}N)^2, \tag{B.55}
$$

with degeneracies (recalling that $\operatorname{Tr} \mathbb{1} = 2^{(N-1)/2}$)

$$
d_l = 2^{(N+1)/2} \frac{\Gamma(l+N)}{l!\,\Gamma(N)} . \tag{B.56}
$$

The calculation proceeds now in close analogy to the scalar case. The identity that allows us to sum the series is in this case

$$
\begin{aligned}
[(l + \tfrac{1}{2}N)^2 + m^2\rho^2]^{n/2} &= \frac{\pi^{1/2}}{\Gamma(-\frac{1}{2}n)} (2m\rho)^{n/2+1/2} \\
&\quad \times \int_0^{\infty} \mathrm{d}t\, \mathrm{e}^{-(l+N/2)t} t^{-n/2-1/2} J_{-n/2-1/2}(m\rho t), \\
&\qquad \operatorname{Re} n > 0 .
\end{aligned}
\tag{B.57}
$$

Proceeding now in precise analogy to the scalar case yields finally

$$
V = (-1)^{\nu+1} \frac{3 \cdot 2^{\nu+1}}{\pi^6 \rho^4} \int_0^{\pi/2} \frac{\mathrm{d}\theta \cos^3\theta \cos 5\theta}{[2 \operatorname{ch}(\frac{1}{2}\pi \tan\theta)]^{2\nu+1}}, \qquad N \text{ odd}, n = 4, m = 0 . \tag{B.58}
$$

As in the scalar cases this integral may be evaluated by residues for low values of ν. This expresses V in terms of η functions where

$$
\eta(s) = \sum_{p=1}^{\infty} \frac{(-1)^{p-1}}{p^s} = \left(1 - \frac{1}{2^{s-1}}\right) \zeta(s); \tag{B.59}
$$

for $\nu = 0, 1$ (i.e. $N = 1, 3$) we find

$$\rho^4 V_1 = -\frac{12\eta(5)}{(2\pi)^6} = -1.895927872 \times 10^{-4}, \tag{B.60}$$

$$\rho^4 V_3 = 3\frac{\eta(5)}{(2\pi)^6} + 360\frac{\eta(7)}{(2\pi)^8} = 1.945058002 \times 10^{-4}. \tag{B.61}$$

Appendix C

THE ASYMPTOTIC FORM OF C FOR LARGE N

Minimally coupled scalars

As $\nu \to \infty$, we make the replacements

$$\left.\begin{matrix} \mathrm{ch}\,(\nu\pi\lambda) \\ \mathrm{sh}\,(\nu\pi\lambda) \end{matrix}\right\} \to \tfrac{1}{2}\,\mathrm{e}^{\nu\pi\lambda}, \tag{C.1}$$

in (B.21) at the cost of an exponentially small error. It will become clear as we proceed that the dominant term in (B.22) is indeed the one with coefficient ν^2. In fact we shall see that

$$C_\nu = -\frac{\nu^2}{8\pi}\int_0^\infty \frac{\mathrm{d}\lambda}{(1+\mathrm{e}^{-\pi\lambda})^{2\nu}}\,\frac{(\lambda^3-3\lambda)}{(\lambda^2+1)^3\pi^3} + \mathrm{O}\left(\frac{\nu}{\log \nu}\right). \tag{C.2}$$

The integrand in (C.2) has its maximum when $\pi\lambda$ is very close to μ where $\mu(\nu)$ is the large root of the equation

$$\frac{\mathrm{e}^{\mu}}{\mu} = \tfrac{2}{3}\nu. \tag{C.3}$$

Specifically

$$\mu = \log[\tfrac{2}{3}\nu \log[\tfrac{2}{3}\nu \log[\tfrac{2}{3}\nu \log[\cdots]]]]. \tag{C.4}$$

It is evident that

$$\mu = \log \tfrac{2}{3}\nu + \log \mu, \tag{C.5}$$

and hence also that

$$\frac{\mu}{\log \frac{2}{3}\nu} \to 1, \tag{C.6}$$

as $\nu \to \infty$.

In (C.2) set

$$\lambda = \frac{\mu x}{\pi}. \tag{C.7}$$

This keeps the maximum near $x=1$ and yields the expression

$$C_\nu=-\frac{\nu^2}{8\pi^2\mu^2}\int_0^\infty \mathrm{d}x \exp\left[-2\nu\log\left(1+\mathrm{e}^{-\mu x}\right)\right]\frac{\left(x^3-\left(3\pi^2/\mu^2\right)x\right)}{\left(x^2+\pi^2/\mu^2\right)^3}+\mathrm{O}(\nu^{-1}\mu^{-1})\,. \tag{C.8}$$

Now

$$\begin{aligned}2\nu\log\left(1+\mathrm{e}^{-\mu x}\right)&=2\nu\,\mathrm{e}^{-\mu x}+\mathrm{O}(\nu\,\mathrm{e}^{-2\mu})\\&=\frac{3}{\mu}\,\mathrm{e}^{-\mu(x-1)}+\mathrm{O}(\nu^{-1}\mu^{-2})\,,\end{aligned} \tag{C.9}$$

in virtue of (B.3).

$$\text{As } \nu\to\infty\,,\qquad \exp\left(-\frac{3}{\mu}\,\mathrm{e}^{-\mu(x-1)}\right)\to\begin{cases}1, & x>1\\ 0, & x<1,\end{cases} \tag{C.10}$$

however the scale over which the function changes from zero to unity is of order $\mu^{-1}\log\mu$. If we, therefore, make the replacement

$$\exp\left(-2\nu\log\left(1+\mathrm{e}^{-\mu x}\right)\right)\to\theta(x-1)\,, \tag{C.11}$$

in (C.8) and retain only the leading term in $1/\mu$ we obtain

$$C_N=-\left(\frac{\nu}{4\pi\mu}\right)^2\left(1+\mathrm{O}(\mu^{\ 1}\log\mu)\right),\qquad m=0\,,\qquad N\text{ odd}\,,\qquad N\to\infty\,. \tag{C.12}$$

Unfortunately this result, apart from showing that C_N eventually becomes negative, is of little practical use owing to the fact that μ increases only slowly with ν.

Fermions

It is much simpler to examine the asymptotic behavior of the Dirac field owing to the fact that (B.53) may be analyzed by means of the method of steepest descent. A short calculation reveals that

$$\begin{gathered}C_\nu=\frac{(-1)^{\nu+1}3}{2^\nu\pi^{13/2}\left(\nu+\frac{1}{2}\right)^{1/2}}\left\{1-\left(\frac{30}{\pi^2}-\frac{1}{8}\right)\frac{1}{\left(\nu+\frac{1}{2}\right)}+\mathrm{O}\left(\frac{1}{\left(\nu+\frac{1}{2}\right)^2}\right)\right\},\\ m=0\,,\qquad N\text{ odd}\,,\qquad N\to\infty\,.\end{gathered} \tag{C.13}$$

Appendix D

We present here a different approach to the evaluation of the quantity

$$C_N^{(0)}=-\frac{1}{32\pi^2}\lim_{n\to4}\Gamma\left(-\tfrac{1}{2}n\right)\sum_l D_l\Lambda_l^{\,n}\,, \tag{D.1}$$

for spheres of odd dimensionality N. Following ref. [25] we write the eigenvalue Λ_l^2 as

$$\Lambda_l^2=l(l+N-1)=L^2-\nu^2\,, \tag{D.2}$$

$$L \equiv l+\nu\,, \qquad \nu \equiv \tfrac{1}{2}(N-1)\,, \tag{D.3}$$

and expand

$$\Lambda_l^n = \sum_{r=0}^{\infty} \frac{\Gamma(r-\frac{1}{2}n)}{\Gamma(-\frac{1}{2}n)r!}\,\nu^{2r}L^{n-2r}\,. \tag{D.4}$$

Also, we write the degeneracy D_l as

$$\begin{aligned} D_l &= \frac{(2l+N-1)(l+N-2)!}{(N-1)!\,l!} \\ &= \frac{2L^2}{(2\nu)!}[L^2-(\nu-1)^2][L^2-(\nu-2)^2]\cdots[L^2-1]\,. \end{aligned} \tag{D.5}$$

Eq. (D.5) can be written as a polynomial in L^2:

$$D_l = \sum_{m=0}^{\nu-i} a_{\nu m} L^{2m+2}\,, \tag{D.6}$$

with $a_{\nu m}$ independent of L (i.e. of l):

$$\begin{aligned} &a_{10}=1\,, && a_{20}=-\tfrac{1}{12}\,, && a_{21}=\tfrac{1}{12}\,, \\ &a_{30}=\tfrac{1}{90}\,, && a_{31}=-\tfrac{1}{72}\,, && a_{32}=\tfrac{1}{360} \qquad \text{etc.} \end{aligned} \tag{D.7}$$

Using (D.6) and (D.5) in (D.1), we have now

$$C_N^{(0)} = -\frac{1}{32\pi^2}\sum_{m=0}^{\nu-1} a_{\nu m} \lim_{n\to 4} \sum_{r=0}^{\infty} \frac{\Gamma(r-\frac{1}{2}n)}{r!}\,\nu^{2r}\zeta(2r-2m-n-2,\,\nu)\,, \tag{D.8}$$

where ζ is the incomplete Riemann zeta function

$$\zeta(z,\nu) = \sum_{L=\nu}^{\infty} L^{-z}\,. \tag{D.9}$$

The factor $\Gamma(-\frac{1}{2}n)$ in (D.1), which blows up for $n \to 4$, has been cancelled by the $\Gamma(-\frac{1}{2}n)$ in the denominator of eq. (D.4). In its place we now have a factor $\Gamma(r-\frac{1}{2}n)$, which blows up for $n \to 4$ only in the cases $r=0$, 1, and 2. To see how this pole is cancelled, note first that because (D.5) vanishes for $L=1, 2, \ldots \nu-1$, the sum over L can be taken to run from 1 to ∞ instead of ν to ∞ in any of the incomplete zeta functions in (D.8), and in particular for the singular terms with $r=0$, 1, and 3, so that in these terms in place of an incomplete zeta function we have the ordinary zeta function $\zeta(2r-2m-n-2)$. But for $r \leq m+2$, the argument $2r-2m-n-2$ approaches a negative even integer as $n \to 4$, where the zeta function has a simple zero that cancels the pole in $\Gamma(r-\frac{1}{2}n)$. Thus there is no singularity here for $n \to 4$.

In evaluating the terms in (D.8) with $r=0, 1, 2$, it is convenient to use the identity [26]

$$\Gamma(\tfrac{1}{2}z)\zeta(z) = \pi^{z-1/2}\,\Gamma\!\left(\frac{1-z}{2}\right)\zeta(1-z)\,, \tag{D.10}$$

which yields

$$\Gamma(r-\tfrac{1}{2}n)\zeta(2r-2m-n-2) \xrightarrow[n\to 4]{} (-)^{m+1}\pi^{2r-2m-13/2}$$
$$\times(3+m-r)!\,\Gamma(\tfrac{7}{2}+m-r)\zeta(7+2m-2r)/(2-r)! \qquad \text{(D.11)}$$

Inserting these terms in (D.8), we have now

$$C_N^{(0)} = -\frac{1}{32\pi^2}\sum_{m=0}^{\nu-1} a_{\nu m}$$
$$\times\left[(-)^{m+1}\pi^{-2m-13/2}\sum_{r=0,1,2}\frac{(\pi\nu)^{2r}(3+m-r)!\,\Gamma(\tfrac{7}{2}+m-r)\zeta(7+2m-2r)}{r!\,(2-r)!}\right.$$
$$\left.+\sum_{r=m+3}^{\infty}\frac{\nu^{2r}}{r(r-1)(r-2)}\zeta(2r-2m-6,\,\nu)\right]. \qquad \text{(D.12)}$$

The convergence of this sum is apparent: the incomplete zeta function $\zeta(2r-2m-6,\ \nu)$ behaves for $r\to\infty$ like $\nu^{-2r+2m+6}$, so the series converges like $\sum 1/r^3$. Eq. (D.12) can therefore be used in numerical calculations of $C_N^{(0)}$.

As a check on our calculation of C_N in appendix B, we have used the more cumbersome series (D.12) to calculate $C_N^{(0)}$ in the case $N=3$ (i.e. $\nu=1$). Here (D.12) becomes

$$C_3^{(0)} = \frac{-1}{32\pi^2}\left[-\pi^{-13/2}\sum_{r=0,1,2}\frac{\pi^{2r}(3-r)!\,\Gamma(\tfrac{7}{2}-r)\zeta(7-2r)}{r!\,(2-r)!}\right.$$
$$\left.+\sum_{r=3}^{\infty}\frac{1}{r(r-1)(r-2)}\zeta(2r-6)\right]. \qquad \text{(D.13)}$$

Since $\zeta(2r-6)$ approaches unity as $r\to\infty$, it is very convenient to improve the convergence of the series in (D.13) by subtracting the elementary series

$$\sum_{r=3}^{\infty}\frac{1}{r(r-1)(r-2)} = \frac{1}{4}.$$

We then have

$$C_3^{(0)} = -\frac{1}{32\pi^2}\left[\frac{-45\zeta(7)}{8\pi^6}-\frac{3\zeta(5)}{2\pi^4}-\frac{\zeta(3)}{4\pi^2}+\tfrac{1}{4}+\sum_{r=3}^{\infty}\frac{1}{r(r-1)(r-2)}\left(\zeta(2r-6)-1\right)\right].$$

To show the speed of convergence of the series, here are the terms up to $r=8$:

$$C_3^{(0)} = -\frac{1}{32\pi^2}[-0.00589976-0.01596762-0.03044846+\tfrac{1}{4}$$
$$-\tfrac{1}{4}+0.02687225+0.00137205+0.00014453$$
$$+0.00001942+0.00000296+\cdots] = 7.568(9)\times 10^{-5}.$$

This agrees with the more precise result derived by the methods of appendix B, shown in table 1.

For *spinors* we have

$$C_N^{(1/2)} = +\frac{1}{8\pi^2} \lim_{n\to 4} \Gamma(-\tfrac{1}{2}n) \sum_{l=0}^{\infty} d_l \lambda_l^n \,. \tag{D.14}$$

The degeneracies and masses derived in appendix B can be written

$$\lambda_l = l + \nu + \tfrac{1}{2}\,, \tag{D.15}$$

$$d_l = \frac{2^{(N+1)/2}}{(N-1)!} \frac{(l+N-1)!}{l!}$$

$$= \frac{2^{\nu+1}}{(2\nu)!} \sum_{m=0}^{\nu} \mathcal{B}_{\nu m} (l+\nu+\tfrac{1}{2})^{2m}\,, \tag{D.16}$$

where again $\nu \equiv \frac{1}{2}(N-1)$, and now

$$\mathcal{B}_{00} = 1\,, \qquad \mathcal{B}_{11} = 1\,, \qquad \mathcal{B}_{10} = -\tfrac{1}{4}\,,$$

$$\mathcal{B}_{22} = 1\,, \qquad \mathcal{B}_{21} = -\tfrac{5}{2}\,, \qquad \mathcal{B}_{20} = \tfrac{9}{16} \cdots . \tag{D.17}$$

Since the sum over m evidently vanishes when $l = -1, -2, \ldots, -(N-1)$, we can let the sum over l run from $l = -\nu$ to ∞ instead of from $l = 0$ to ∞; setting $L = l + \nu$, this gives

$$C_N^{(1/2)} = \frac{2^\nu}{4\pi^2(2\nu)!} \sum_{m=0}^{\nu} \mathcal{B}_{\nu m} \lim_{n\to 4} \Gamma(-\tfrac{1}{2}n) \sum_{L=0}^{\infty} (L+\tfrac{1}{2})^{n+2m} . \tag{D.18}$$

The sum over L can easily be expressed in terms of ordinary Riemann zeta functions

$$\sum_{L=0}^{\infty} (L+\tfrac{1}{2})^{n+2m} = -(1-2^{-n-2m})\zeta(-n-2m)\,.$$

Using (D.10) again, the limit $n \to 4$ gives

$$\Gamma(-\tfrac{1}{2}n)\zeta(-n-2m) \xrightarrow[n\to 4]{} \tfrac{1}{2}(-)^m \pi^{-2m-9/2}(2+m)!\,\Gamma(\tfrac{5}{2}+m)\zeta(5+2m)\,.$$

Putting this all together, we find

$$C_N^{(1/2)} = \frac{-2^\nu}{8\pi^2(2\nu)!} \sum_{m=0}^{\nu} (1-2^{-4-2m})\mathcal{B}_{\nu m}(-)^m \pi^{-2m-9/2}(2+m)!\,\Gamma(\tfrac{5}{2}+m)\zeta(5+2m)\,. \tag{D.19}$$

For instance, in the case $N = 3$ $(\nu = 1)$, this gives simply

$$C_3^{(1/2)} = \frac{45\zeta(5)}{1024\pi^6} + \frac{2835\zeta(7)}{2048\pi^8} = 1.9451 \times 10^{-4}\,.$$

This reproduces the value derived in appendix B, shown in table 3. At least for fermions, the present method seems an especially convenient way of obtaining expressions for C_N in closed form.

Appendix E

MEAN SQUARE O($N+1$) GENERATORS

In this appendix we calculate Q_l^2 and q_l^2, the eigenvalues of the squared O($N+1$) generators, averaged over the eigenmodes of the laplacian and Dirac operators on a sphere S^N of radius ρ with eigenvalues $M_l^2 = l(l+N-1)/\rho^2$ and $m_l = (l+\frac{1}{2}N)/\rho$, respectively. These mean eigenvalues are the same for all O($N+1$) generators, so they can be expressed as the eigenvalues $\boldsymbol{J}_l^2$ of the O($N+1$) Casimir operators, which are the same for all eigenmodes of definite mass, divided by the number $\frac{1}{2}N(N+1)$ of O($N+1$) generators.

For spin-0, the O($N+1$) Casimir operator is trivially related to the laplacian. Embedding S^N in an ($N+1$)-dimensional euclidean space with coordinates $Z^i(i=1,2,\dots,N+1)$, we have SO($N+1$) generators

$$J_{ij} = L_{ij} \equiv -i\left(Z^i \frac{\partial}{\partial Z^j} - Z^j \frac{\partial}{\partial Z^i}\right), \tag{E.1}$$

so the Casimir operator is

$$\boldsymbol{J}^2 \equiv \tfrac{1}{2}\sum_{ij} J_{ij}^2 = -\rho^2\nabla^2 + \rho^2 \frac{\partial^2}{\partial\rho^2} + N\rho\frac{\partial}{\partial\rho}, \tag{E.2}$$

where

$$\rho^2 \equiv \sum_i (Z^i)^2 \tag{E.3}$$

$$\nabla^2 \equiv \sum_i \frac{\partial^2}{\partial Z^{i2}}. \tag{E.4}$$

Acting on a function of the unit vector Z^i/ρ, for which the laplacian ∇^2 has eigenvalue $-M_l^2$, the Casimir operator has eigenvalues

$$\boldsymbol{J}_l^2 = \rho^2 M_l^2 = l(l+N-1), \tag{E.5}$$

and therefore

$$Q_l^2 = \frac{\rho^2 M_l^2}{\frac{1}{2}N(N+1)} = \frac{l(l+N-1)}{\frac{1}{2}N(N+1)}. \tag{E.6}$$

An immediate consequence of the relation between Q_l^2 and M_l^2 is that, comparing (56) and (58),

$$D_N^{(0)} = -\frac{4}{N(N+1)} E_N^{(0)}. \tag{E.7}$$

For spin-$\frac{1}{2}$ the $\mathrm{O}(N+1)$ generators involve an intrinsic as well as an orbital term

$$J_{ij} = L_{ij} + S_{ij}\,, \tag{E.8}$$

where L_{ij} is as before

$$L_{ij} = -i\left(Z^i \frac{\partial}{\partial Z^j} - Z^j \frac{\partial}{\partial Z^i}\right),$$

and now

$$S_{ij} = -\tfrac{1}{4} i [\Gamma_i, \Gamma_j]\,, \tag{E.9}$$

where Γ_i are the gamma matrices in $(N+1)$-dimensional euclidean space

$$\{\Gamma_i, \Gamma_j\} = 2\delta_{ij}\,. \tag{E.10}$$

The Casimir operator is then

$$\boldsymbol{J}^2 \equiv \tfrac{1}{2} \sum_{ij} (L_{ij} + S_{ij})^2\,. \tag{E.11}$$

Following the approach of ref. [27], an elementary manipulation of the commutation relations for Z^i and $\partial/\partial Z^j$ and the anticommutation relations for Γ^i and Γ^j shows that

$$(\boldsymbol{J}^2 - \boldsymbol{L}^2 - \boldsymbol{S}^2)^2 - (N-1)(\boldsymbol{J}^2 - \boldsymbol{L}^2 - \boldsymbol{S}^2) - \boldsymbol{L}^2 = 0\,, \tag{E.12}$$

with $\boldsymbol{J}^2$, $\boldsymbol{L}^2$, and $\boldsymbol{S}^2$ interpreted as

$$\boldsymbol{A}^2 \equiv \tfrac{1}{2} \sum_{ij} (A_{ij})^2\,.$$

Also,

$$\boldsymbol{S}^2 = \tfrac{1}{8} N(N+1)\,, \tag{E.13}$$

and, acting on the kth spherical harmonic on S^N,

$$\boldsymbol{L}^2 = k(k+N-1)\,. \tag{E.14}$$

It follows that the direct product of a Dirac spinor and a spherical harmonic of order k is a mixture of eigenmodes of $\boldsymbol{J}^2$ with eigenvalues given by the two roots of (E.12):

$$\boldsymbol{J}^2 = \begin{cases} k^2 + kN + \tfrac{1}{8} N(N+1) \\ (k-1)^2 + (k-1)N + \tfrac{1}{8} N(N+1)\,. \end{cases} \tag{E.15}$$

In other words, the eigenvalues of $\boldsymbol{J}^2$ take the form

$$\boldsymbol{J}_l^2 = l^2 + lN + \tfrac{1}{8} N(N+1)\,, \qquad l = 0, 1, \ldots, \tag{E.16}$$

with eigenmodes consisting of mixtures of spherical harmonics with $k = l$ and

$k=l+1$. The eigenmodes of the Dirac operator are of course also eigenmodes of $\boldsymbol{J}^2$, and the integer l can be identified as the integer used to label the eigenvalues $m_l=(l+\frac{1}{2}N)/\rho$ of the Dirac operator. In accordance with our earlier remarks, the mean square $\mathrm{O}(N+1)$ generator is given by

$$q_l^2=\boldsymbol{J}_l^2/\tfrac{1}{2}N(N+1)=\frac{l^2+lN+\frac{1}{8}N(N+1)}{\frac{1}{2}N(N+1)}\,. \tag{E.17}$$

We can easily check this for the familiar case of O(3), i.e. $N=2$. In this case (E.5) gives the familiar result for spinless fields, $\boldsymbol{J}_l^2=l(l+1)$. Also, for spin-$\frac{1}{2}$ a spherical harmonic of order k will have total spin $j=k\pm\frac{1}{2}$, and hence the Casimir operator

$$\boldsymbol{J}^2=j(j+1)=(k\pm\tfrac{1}{2})(k\pm\tfrac{1}{2}+1)$$

in agreement with (E.15) for $N=2$.

Appendix F

EVALUATION OF THE COEFFICIENTS D_N AND E_N

We shall calculate, in this appendix, the coefficients that renormalize the gravitational constant and the Yang–Mills coupling.

Scalar fields

For scalar fields it is sufficient to compute $E_N^{(0)}$ since $D_N^{(0)}$ and $E_N^{(0)}$ are related by eq. (E.7). From eq. (58) we have

$$\frac{E_N^{(0)}}{\rho^2}=\frac{1}{192\pi^2}\lim_{n\to 4}\Gamma(1-\tfrac{1}{2}n)\sum_{l=0}^{\infty}D_l(M_l^2)^{n/2-1}\,, \tag{F.1}$$

while $V(m^2)$, the effective potential for a scalar field of mass m, is given by (B.7) in the form

$$V(m^2)=-\frac{1}{32\pi^2}\lim_{n\to 4}\Gamma(-\tfrac{1}{2}n)\sum_{l=0}^{\infty}D_l(M_l^2+m^2)^{n/2}\,. \tag{F.2}$$

Comparing (F.1) and (F.2) and setting $n=4$ we obtain

$$\frac{E_N}{\rho^2}=\tfrac{1}{6}V'(0)\,,\qquad \text{scalars, } n=4\,, \tag{F.3}$$

where a prime denotes $\partial/\partial m^2$.

We shall calculate $V'(m^2)$ from (B.15). Recalling that

$$b^2=\nu^2-m^2\rho^2\,, \tag{F.4}$$

we see that the m dependence of V is contained in the quantity

$$(bt)^{7/2}I_{-7/2}(bt)\,.$$

Setting $z = bt$ we have

$$\frac{\partial}{\partial m^2}(z^{7/2}I_{-7/2}(z)) = -\tfrac{1}{2}\rho^2 t^2 \frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}(z^{7/2}I_{-7/2}(z))$$

$$= -\tfrac{1}{2}\rho^2 t^2 z^{5/2} I_{-5/2}(z)\,, \tag{F.5}$$

where, in the last equality, we have used the identity

$$\frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}(z^{-\nu}I_\nu(z)) = z^{-(\nu+1)}I_{\nu+1}(z)\,. \tag{F.6}$$

Noting also that

$$z^{5/2}I_{-5/2}(z) = \sqrt{\frac{2}{\pi}}\{-3z\,\mathrm{sh}\,z + (3+z^2)\,\mathrm{ch}\,z\}\,, \tag{F.7}$$

we obtain E_N in the form

$$E_N^{(0)} = \frac{1}{96\pi^2\nu}\int_c \frac{\mathrm{d}t}{t^4(2\,\mathrm{sh}\,\frac{1}{2}t)^{2\nu}}[-3\nu t\,\mathrm{sh}\,(\nu t) + (3+\nu^2 t^2)\,\mathrm{ch}\,(\nu t)]\,. \tag{F.8}$$

As in previous cases evaluation by residues yields exact results. Thus, by way of example, we find

$$E_3^{(0)} = \frac{1}{24}\left(\frac{5\zeta(3)}{(2\pi)^4} - \frac{12\zeta(5)}{(2\pi)^6}\right)$$

$$= 1.522542\times 10^{-4}\,. \tag{F.9}$$

To facilitate numerical computation for general values of N, we again set $t = \pi i + \pi\tan\theta$ and obtain as our final expression

$$E_N^{(0)} = \frac{1}{48\pi^2\nu}\int_0^{\pi/2}\frac{\mathrm{d}\theta}{[2\,\mathrm{ch}\,(\frac{1}{2}\pi\tan\theta)]^{2\nu}}\left\{3\nu\cos\theta\sin 3\theta\,\mathrm{sh}\,(\nu\pi\tan\theta)\right.$$

$$\left. + \left(\frac{3}{\pi}\cos^2\theta\cos 4\theta - \nu^2\pi\cos 2\theta\right)\mathrm{ch}\,(\nu\pi\tan\theta)\right\}. \tag{F.10}$$

Numerical values of this integral are used (along with the values of $C_N^{(0)}$ obtained in appendix B) to calculate the values of the coefficients $\beta_N^{(0)}$ and $\gamma_N^{(0)}$ presented in table 4.

Fermions

For massless fermions we have

$$\frac{E_N^{(1/2)}}{\rho^2} = \frac{1}{96\pi^2}\lim_{n\to 4}\Gamma(1-\tfrac{1}{2}n)\sum_{l=0}^{\infty} d_l(m_l^2)^{n/2-1}\,, \tag{F.11}$$

$$D_N^{(1/2)} = -\frac{1}{192\pi^2}\,\frac{20\nu^2+28\nu+5}{(\nu+1)^2}\,\lim_{n\to 4}\Gamma\left(2-\frac{n}{2}\right)\sum_l d_l(m_l^2)^{n/2-2}$$
$$\times\frac{[l^2+lN+\frac{1}{8}N(N+1)]}{\frac{1}{2}N(N+1)}\,. \tag{F.12}$$

As in the scalar case we proceed by relating $E_N^{(1/2)}$ and $D_N^{(1/2)}$ to the derivative of $V(m^2)$,. which is given for spin-$\frac{1}{2}$ by (B.54) as

$$V(m^2) = \frac{1}{8\pi^2}\lim_{n\to 4}\Gamma(-\tfrac{1}{2}n)\sum_{l=0}^{\infty} d_l(m_l^2+m^2)^{n/2}\,. \tag{F.13}$$

We see that

$$\frac{E_N}{\rho^2} = -\tfrac{1}{12}V'(0)\,, \qquad \text{fermions, } n=4\,, \tag{F.14}$$

and by noting that we may write

$$l^2+lN+\tfrac{1}{8}N(N+1) = m_l^2\rho^2-\tfrac{1}{8}N(N-1)\,, \tag{F.15}$$

we find

$$D_N^{(1/2)} = \frac{20\nu^2+28\nu+5}{(\nu+1)^2}\left\{\frac{E_N^{(1/2)}}{N(N+1)}+\frac{1}{96}\left(\frac{N-1}{N+1}\right)V''(0)\right\}, \qquad \text{fermions,} \qquad n=4\,. \tag{F.16}$$

Proceeding as in the scalar case we make use of a representation that follows from (B.57)

$$V(m^2) = \frac{2^{\nu-1/2}}{\pi^{3/2}\rho^4}\int_C \frac{\mathrm{d}t}{t^5}\,\frac{(mt\rho)^{5/2}J_{-5/2}(mt\rho)}{(2\,\mathrm{sh}\,\frac{1}{2}t)^{2\nu+1}}\,. \tag{F.17}$$

The derivatives of V are readily calculated in virtue of the relations

$$\left(\frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}\right)^k (z^{-\nu}J_\nu(z)) = (-1)^k z^{-(\nu+k)}J_{\nu+k}(z)\,, \tag{F.18}$$

and

$$\lim_{z\to 0} z^{-\nu}J_\nu(z) = \frac{2^{-\nu}}{\Gamma(1+\nu)}\,. \tag{F.19}$$

Thus we find

$$E_N^{(1/2)} = \frac{(-1)^{\nu+1}2^\nu}{12\pi^4}\int_0^{\pi/2}\frac{\mathrm{d}\theta\cos\theta\cos 3\theta}{[2\,\mathrm{ch}\,(\frac{1}{2}\pi\tan\theta)]^{2\nu+1}}\,, \qquad \text{fermions, } n=4\,, \tag{F.20}$$

$$V''(0) = \frac{(-2)^{\nu-1}}{\pi^2}\int_0^{\pi/2}\frac{\mathrm{d}\theta}{[2\,\mathrm{ch}\,(\frac{1}{2}\pi\tan\theta)]^{2\nu+1}}\,, \qquad \text{fermions, } n=4\,, \tag{F.21}$$

from which $D_N^{(1/2)}$ is obtained using (F.16). Numerical values of these integrals for $D_N^{(1/2)}$ and $E_N^{(1/2)}$ are used (along with the values of $C_N^{(1/2)}$ obtained in appendix B) to calculate the values of the coefficients $\beta_N^{(1/2)}$ and $\gamma_N^{(1/2)}$ presented in table 5.

As a check, we have also calculated $D_N^{(j)}$ and $E_N^{(j)}$ for $j=0$ and $\frac{1}{2}$ and low N by the summation method of appendix B. The results confirm those obtained in this appendix.

References

[1] Th. Kaluza, Sitz. Preuss. Akad. Wiss. K1 (1921) 966;
O. Klein, Nature 118 (1926) 516; Z. Phys. 37 (1926) 895;
H. Mandel, Z. Phys. 39 (1926) 136;
A. Einstein and P. Bergmann, Ann. of Math. 39 (1938) 683

[2] B. DeWitt, in Relativity, groups, and topology (Gordon and Breach, New York and London, 1964) p. 725;
J. Scherk and J.H. Schwarz, Phys. Lett. 57B (1975) 463;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
E. Witten, Nucl. Phys. B186 (1981) 412;
E. Cremmer, in Supergravity 81, eds. S. Ferrara and J. Taylor (Cambridge University Press, 1982);
A. Salam and J. Strathdee, Ann. of Phys. 141 (1982) 316;
M.J. Duff and D. Toms, in Unification of fundamental interactions 2, eds. S. Ferrara and J. Ellis (Plenum, 1982);
C. Wetterich, Phys. Lett. 110B (1982) 379,
F. Englert, Phys. Lett. 119B (1982) 339

[3] E. Cremmer and J. Scherk, Nucl. Phys. B108 (1976) 408;
Z. Horvath, L. Palla, E. Cremmer and J. Scherk, Nucl. Phys. B127 (1977) 57;
J.F. Luciani, Nucl. Phys. B135 (1978), 111; P.G.O. Freund and M. Rubin, Phys. Lett. 97B (1980) 233;
M. Duff, in Supergravity 81, eds. S. Ferrara and J. Taylor (Cambridge University Press, 1982);
S. Randjbar-Daemi, A. Salam and J. Strathdee, ICTP preprints IC/82/208 (1982), IC/83/6 (1983);
M.J. Duff and C.N. Pope, Imperial College preprint ICTP/82–83/7 (1983);
M.J. Duff, B. Nilsson, and C. Pope, Texas preprint UTTG-5 (1983)

[4] S. Weinberg, Phys. Lett. 125B (1983) 265

[5] T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141;
D. Pollard, ICTP preprint 82–83/11;
M.A. Rubin and B.D. Roth, Texas preprint UTTG-3 (1983), Phys. Lett. B, to appear;
T. Appelquist, A. Chodos and E. Myers, Yale preprint YTP 83–04 (1983);
T. Appelquist and A. Chodos, Yale preprint YTP 83-05 (1983)

[6] J.B. Hartle and G.T. Horowitz, Phys. Rev. D24 (1981) 257;
Also see E. Tomboulis, Phys. Lett. 70B (1977) 361; 97B (1980) 77

[7] M.J. Duff and D. Toms, in Unification of fundamental interactions, eds. S. Ferrara and J. Ellis (Plenum, 1982)

[8] S. Coleman and E. Weinberg, Phys. Rev. D7 (1973) 1888;
S. Weinberg, Phys. Rev. D7 (1973) 1068

[9] A. Erdelyi, W. Magnus, F. Oberhettinger and F.R. Tricomi, Higher transcendental functions (New York, McGraw-Hill, 1953)

[10] B.S. DeWitt, *Dynamical theory of groups and fields* (Gordon and Breach, 1964);
D.J. Toms, University of Wisconsin at Milwaukee preprint;
M. Awada and D.J. Toms, University of Wisconsin preprint.

[11] S. Weinberg, in Gravitational theories since Einstein, eds. S. Hawking and W. Israel (Cambridge University Press, 1978), sect. 16.7

[12] A. Zee, Phys. Rev. D23 (1981) 858

[13] K. Symanzik, Comm. Math. Phys. 16 (1970) 48;
S. Coleman, in Laws of hadronic matter, ed. A. Zichichi (Academic Press, NY, 1975) 139

[14] D. Page, paper in preparation

[15] H. Weyl, Rend. Circ. Mat. Palermo 31 (1915) 1

[16] M.A. Rubin and B.D. Roth, Nucl. Phys. B226 (1983) 444

[17] S. Weinberg, Gravitation and cosmology (Wiley, New York, 1972) Chap. 13

[18] E. Witten, private communication

[19] M.J. Duff, C.N. Pope and B.E.W. Nilsson, private communication

[20] A. Zee, in Proc. 4th Summer Institute on grand unified theories and related topics, eds. M. Konuma and T. Maskawa (World Scientific, Singapore, 1981) appendix VI-5

[21] D.G. Boulware and L.S. Brown, Ann. of Phys. 138 (1982) 392

[22] S. Helgason, Differential geometry and symmetric spaces (Academic Press, New York, 1962)

[23] M.E. Peskin, Nucl. Phys. B175 (1981) 197

[24] P. Candelas and D.J. Raine, Phys. Rev. D12 (1975) 965

[25] S.M. Christensen and M.J. Duff, Nucl. Phys. B170 (1980) 480

[26] I.S. Gradshteyn and I.M. Ryzik, Table of integrals, series, and products (Academic Press, New York, 1980)

[27] I.T. Drummond and G.M. Shore, Ann. of Phys. 117 (1979) 89

Spontaneous compaction in Kaluza–Klein models and the Casimir effect

N. A. Voronov and Ya. I. Kogan
Institute of Theoretical and Experimental Physics

(Submitted 14 July 1983)
Pis'ma Zh. Eksp. Teor. Fiz. **38**, No. 5, 262–265 (10 September 1983)

A particular mechanism for spontaneous compaction in Kaluza–Klein models is analyzed. The compaction in this case results from the quantum vacuum expectation value of the energy-momentum tensor of the matter fields. It is shown that the space $M^4 \times S^d$ is a solution of the effective Einstein equations.

PACS numbers: 12.10. – g

Models of the Kaluza–Klein type[1] have been discussed extensively in recent years in connection with the effort to unify gravitation and gauge fields[2] and to construct expanded supergravity models.[3] Here it is necessary to compact the original $(4+d)$-dimensional space in order to make the ground state—the direct product of the non-compact four-dimensional space-time and the compact d-dimensional inner space—a solution of the corresponding Einstein equations. All the compaction mechanisms that have been discussed[4] use nontrivial classical solutions of the Einstein equations with auxiliary matter fields.

We wish to propose another possible compaction mechanism, according to which the energy-momentum tensor on the right side of the Einstein equations arises not from the classical solutions but as a result of incorporating single-loop corrections associated with the compactness of the d-dimensional space. This is none other than the well-known Casimir effect.[5] Appelquist and Chodos[6] have studied the Casimir effect in the case of the compaction of $M^4 \times S^1$ in connection with questions related to the stability of the vacuum.

The Einstein equations in a $(4+d)$-dimensional space are

$$R^{(d)}_{AB} - \frac{1}{2} g_{AB} R^{(d)} = \kappa (T_{AB} + g_{AB} \Lambda), \tag{1}$$

where κ is the $(d+4)$-dimensional gravitational constant, which is related to the Newtonian constant κ_N by $\kappa = \kappa_N V(S^d)$; $V(S^d)$ is the volume of S^d; $A, B = 0, 1, \ldots, 4+d-1$; T_{AB} is the vacuum expectation value of the energy-momentum tensor; and the value of the Λ term is determined from the vanishing of the curvature of the four-dimensional space M^4, If we wish the solution of (1) to be an $M^4 \times S^d$ space, then we have

$$g_{AB} = \left(\begin{array}{c|c} \eta_{\mu\nu} & \\ \hline & g_{ab}(\theta_i) \end{array}\right); \quad R_{AB} = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \frac{R}{d} g_{ab} \end{array}\right), \tag{2}$$

where R is the scalar curvature of S^d, $R = d(d-1)/r^2$; r is the radius of the S^d sphere;

and θ_i are the coordinates on the sphere. Substituting (2) into the left side of (1), we find

$$\kappa (T_{AB} + g_{AB} \Lambda) = \left(\begin{array}{c|c} -\dfrac{d(d-1)}{2r^2} \eta_{\mu\upsilon} & 0 \\ \hline 0 & -\dfrac{(d-1)(d-2)}{2r^2} g_{ab} \end{array} \right) . \tag{3}$$

If (3) is to have a solution, the conditions $\langle 0|T_{\mu\upsilon}|0\rangle = T_1 \eta_{\mu\upsilon}$, $T_{a\mu} = 0$, $\langle 0|T_{ab}|0\rangle = T_2 g_{ab}$, must hold. These conditions naturally must hold for the vacuum expectation value of the energy-momentum tensor in view of the symmetry of the spaces M^4 and S^d. From (3) we find

$$\kappa (T_1 + \Lambda) = -\frac{d(d-1)}{2r^2} \quad ; \quad \kappa (T_2 + \Lambda) = -\frac{(d-1)(d-2)}{2r^2} \quad . \tag{4}$$

For a conformally invariant coupling of matter with gravity we have the relation $4T_1 + dT_2 = 0$, which is related to the trace-free nature of T_{AB}. This assertion is valid if there are no anomalies[1] in T^A_A (Ref. 7). In general, we would have $g^{AB}(x)\langle 0|T_{AB}(x)|0\rangle = A(x)$, where $A(x)$ is expressed in terms of R^2_{ABCD} and R^2_{AB}. It is important to note that for the maximally symmetric spaces, including S^d, these quantities are actually independent of x, so that $\langle 0|T^B_B|0\rangle = A$ is a constant, determined exclusively by the sphere radius r and the nature of the matter fields. For simplicity, we set $A = 0$ at this point, although all the results can easily be generalized to the case $A \neq 0$ and will be discussed in a separate detailed paper.

Since $T_1 = T_{00}$ is the energy density in the $d + 4$ space, it is given by

$$T_1 = \frac{1}{2V(S^d)} \left[\sum_i \int \frac{d^3k}{(2\pi)^3} N_i \sqrt{\mathbf{k}^2 + \frac{m_i^2}{r^2}} \right]_{\text{reg}} , \tag{5}$$

where m_i^2 are the eigenvalues of the operator of the corresponding equation of motion of the matter field, which is bounded on the compact subspace S^d, and N_i is the number of states with the given m_i. A different expression for T_1 was used in Ref. 6:

$$T_1 = -\frac{i}{2V(S^d)} \left[\sum_i \int \frac{d^4k}{(2\pi)^4} N_i \ln \left(k^2 + \frac{m_i^2}{r^2}\right) \right]_{\text{reg}} , \quad k^2 = \mathbf{k}^2 - k_0^2 .$$

It is a simple matter to show that these two representations lead to the same result for the regularized $\langle 0|T_{AB}|0\rangle$. A calculation of the Casimir effect does not depend[8] on the choice of a regularization method for (5), and it leads to an expression of the type $T_1 = C/r^{4+d}$. The constant C can be calculated unambiguously; it depends on the particular choice of matter fields. From (4) we then find expressions for that value of Λ term, which provides the required compaction of $M^4 \times S^d$ and for the compaction radius r_0:

$$r_0^{2+d} = -\kappa C \frac{(d+4)}{d(d-1)}; \qquad \Lambda = -\frac{d(d-1)(d+2)}{2r_0^2(d+4)} \tag{6}$$

It follows from (6) that for our purposes we must have $C<0$.

Up to this point, we have been talking about the energy-momentum tensor of matter fields. Appelquist and Chodos[6] have calculated an effective potential (i.e., the component T_{00}) which results from quantum fluctuations in the gravitational field above a given background metric. It may be that a tensor structure of the type in (3) arises in this case; this circumstance would mean a purely dynamic compaction of the space in the absence of additional matter fields. However, the validity of this suggestion must be looked at more carefully (R^2 are the contraterms in a single loop).

This compaction mechanism does not work for the $M^4\times S^1$ model because of the zero curvature of M^4 and S^1. The first nontrivial dimensionality is (4 + 2).

Let us briefly[2)] discuss the calculation of the Casimir effect for the model of a compaction of $M^4\times S^2$, in which the matter fields are a massless scalar field and a massless spinor field. The boson spectrum in this case is well known:

$$m_i^2 = i(i+1); \qquad N_i = 2i+1, \qquad i = 0, 1, \ldots, \tag{7}$$

where m_i and N_i are the same as in (5).

The problem of calculating the fermion spectrum is less trivial, because of the spin-gravity interaction.

$$m_f^2 = f^2; \qquad N_f = 4f \qquad f = 1, 2, \ldots \tag{8}$$

Knowing (7) and (8), and using the standard regularization methods,[8] we can evaluate (5) for a specific model. Here we must take into account the circumstances that the contribution of the fermions to the energy of the vacuum enters with a minus sign. We finally find

$$C_b = -1.1\times 10^{-4}; \qquad C_f = 7\times 10^{-4}. \tag{9}$$

We sincerely thank M. B. Voloshin, L. B. Okun', K. A. Ter-Martirosyan, and M. A. Shifman. We thank in particular I. Yu. Kobzarev for many useful discussions.

After this paper had been submitted for publication, we learned of the preprint[9] by Candelas and Weinberg, who discussed some corresponding questions. We wish to thank D. Gross and V. I. Zakharov for bringing this preprint to our attention.

1) We wish to thank V. N. Gribov and M. A. Shifman for a useful and stimulating discussion of this question.

2) The detailed calculations will be published elsewhere.

---

[1] Th. Kaluza, Sitz, Preuss, Akad. Wiss. **K1**, 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926).
[2] B. DeWitt, in: Relativity, Groups and Topology, New York, London, 1964; Y. M. Cho and P. G. O. Freund, Phys. Rev. **D12**, 1711 (1975).
[3] E. Cremmer, B. Julia, and J. Scherk, Phys. Lett. **76B**, 409 (1978); P. van Nieuwenhuizen, Phys. Reports **68**, 189 (1981).
[4] E. Cremmer and J. Scherk, Nucl. Phys. **B108**, 409 (1976); **B118**, 61 (1977); J. F. Luciani, Nucl. Phys. **B135**, 111 (1978); P. G. O. Freund and M. Rubin, Phys. Lett. **97B**, 233 (1980); D. V. Volkov and V. I. Tkach, Teor. Mat. Fiz. **51**, 171 (1982); P. van Nieuwenhuizen, Preprint TH. 3497, CERN, 1983.
[5] H. B. G. Casimir, Proc. Kon. Nederl. Acad. Wet. **51**, 793 (1948).
[6] T. Appelquist and A. Chodos, Phys. Rev. Lett. **50**, 141 (1983); Yale Preprint YTP 83-05, 1983.
[7] K. Fujikawa, Phys. Rev. Lett. **44**, 1733 (1980); M. J. Duff, Preprint TH. 3232, CERN, 1982.
[8] A. A. Grub, S. G. Mamaev, and V. M. Mostepanenko, Kvantovye éffekty v intensivnykh vneshnikh polyakh (Quantum Effects in Intense External Fields), Atomizdat, Moscow, 1980 (see also the bibliography in this book).
[9] P. Candelas and S. Weinberg, Texas Preprint UTTG-6-83.

Translated by Dave Parsons
Edited by S. J. Amoretty

ONE-LOOP EFFECTS ON THE ROUND SEVEN-SPHERE

G.W. GIBBONS

DAMTP, University of Cambridge, UK

and

H. NICOLAI

CERN, Geneva, Switzerland

Received 24 February 1984
Revised manuscript received 30 April 1984

We examine the quantum fluctuations of the massive tower of states arising in the spontaneous compactification of eleven-dimensional supergravity on the round seven-sphere and show that the Casimir energy and the one-loop β function vanish "floor-by-floor". The Euler number divergence, in contrast, is found to vanish only if all higher modes are taken into account.

It has recently been shown that the eleven-dimensional supergravity theory [1] may give rise to effective four-dimensional theories via the mechanism of spontaneous compactification (for recent reviews, see refs. [2,3]). This means that there exist solutions of the classical equations of motion in which the eleven-dimensional spacetime decomposes into the product of four-dimensional anti-de Sitter space and a seven-dimensional, compact, internal space. However from the point of view of the quantum theory this background should be regarded as a sort of "mean field" approximation. It is therefore important to examine the quantum fluctuations around these classical backgrounds. In this letter we present a calculation of the one-loop corrections of the Freund–Rubin parameter for the compactification of the eleven-dimensional supergravity on the round seven-sphere [4,5]. This parameter is related to the size of the internal manifold and thus to the gauge-coupling constant of the four-dimensional theory at the classical level. Therefore our calculation determines the contribution to its β-function from the infinite lower of massive states arising from the higher modes on the seven-sphere. Another quantity of interest is the Casimir energy of these massive states. The formal sum of the Casimir energies of all the states on the round seven-sphere equals the ground-state energy which must vanish if the ground state is supersymmetric. The central result of this paper is that, for the round seven-sphere. both the β-function and the Casimir energy vanish. In fact, there occur more cancellations than are actually required since the various contributions cancel "floor by floor". From these results it follows that the $N = 8$ supersymmetry of the classical ground state [5] is maintained to lowest non-trivial order in the quantized theory, and therefore supersymmetry is not broken by quantum corrections at this order. The relation between the SO(8) gauge coupling constant and the cosmological constant is preserved at one loop. We also consider the contributions to the Euler number divergence. In contrast to the β-function and Casimir energies each floor contributes a non-vanishing amount but their total contribution vanishes.

So far the β-function and Casimir energy have been calculated only for the massless modes. At the linearised level the truncation of the eleven-dimensional theory on the round seven-sphere to the massless supermultiplet coincides with the gauged $N = 8$ supergravity in four dimensions [6,5]. The renormalization of the cosmological constant, which is proportional to the gauge coupling constant, was calculated for all gauged supergravity theories in four dimensions by an explicit computation of the coefficient of the relevant counter-term [7]. For $N > 4$ the coefficient was found to vanish. Later this result was rederived by Curtright [8] who observed that the contribution of a particle of helicity s is given by

$$\beta(g) = (-1)^{2s}(96\pi^2)^{-1}g^3C(1 - 12s^2), \tag{1}$$

where C is the Casimir of the corresponding representation, i.e. the eigenvalue of the Casimir operator times the dimension of the representation. The β-function is then seen to vanish by spin-sum rules. It is interesting to note that formula (1), when applied to the massive tower of states arising in the compactification, gives *negative* contributions for *charged* gravitons and spin one fields, whereas in ordinary four-dimensional gauge theories, negative contributions can only come from self-interacting spin-one fields.

The Casimir energies were calculated by Allen and Davis [9] who used ζ-function regularization of the divergent sum of the energy eigenvalues. Again spin-sum rules imply that the Casimir energy vanishes for $N > 4$ and that the ζ-function has no poles indicating the absence of one-loop divergences. It should be pointed out that the relation of the energy ζ-function regularization to the ζ-function regularization of functional determinants has only been fully worked out in the case when the metric component g_{00} is independent of space and time and $g_{01} = 0$ (the "ultrastatic" case [10]). Its use in anti-de Sitter space is thus somewhat of an extrapolation from known results. For some comments on this problem see ref. [11].

To proceed we recall that the spectrum on the round seven-sphere has recently been determined completely ([3,12–14], for the fermions see also ref. [15]). The results are summarized in table 1. We use Dynkin labels to designate irreducible representations of SO(8) [‡1]. In the first column of table 1 we give the spin and in the last column we show the associated energy labels. For each n table 1 represents an irreducible representation of Osp (8|4). Thus n labels the "floors" of the infinite tower of massive states. The restrictions on n are such that the Dynkin labels only contain non-negative integers. In particular there

Table 1
The spectrum of supergravity on the seven-sphere. The states marked by an asterisk constitute the $N = 8$ singleton multiplet when $n = -1$.

Spin	SO(8) content		E_0
2^+	$(n\,0\,0\,0)$	$n \geqslant 0$	$3 + n/2$
$3/2^{(1)}$	$(n\,0\,0\,1)$	$n \geqslant 0$	$5/1 + n/2$
$3/2^{(2)}$	$(n-1\,0\,1\,0)$	$n \geqslant 1$	$7/2 + n/2$
$1^{-(1)}$	$(n\,1\,0\,0)$	$n \geqslant 0$	$2 + n/2$
1^+	$(n-1\,0\,1\,1)$	$n \geqslant 1$	$3 + n/2$
$1^{(-2)}$	$(n-2\,1\,0\,0)$	$n \geqslant 2$	$4 + n/2$
$1/2^{(1)}$	$(n+1\,0\,1\,0)$ *	$n \geqslant 0$	$3/2 + n/2$
$1/2^{(2)}$	$(n-1\,1\,1\,0)$	$n \geqslant 1$	$5/2 + n/2$
$1/2^{(3)}$	$(n-2\,1\,0\,1)$	$n \geqslant 2$	$7/2 + n/2$
$1/2^{(4)}$	$(n-2\,0\,0\,1)$	$n \geqslant 2$	$9/2 + n/2$
$0^{+(1)}$	$(n+2\,0\,0\,0)$ *	$n \geqslant 0$	$1 + n/2$
$0^{-(1)}$	$(n\,0\,2\,0)$	$n \geqslant 0$	$2 + n/2$
$0^{+(2)}$	$(n-2\,2\,0\,0)$	$n \geqslant 2$	$3 + n/2$
$0^{-(2)}$	$(n-2\,0\,0\,2)$	$n \geqslant 2$	$4 + n/2$
$0^{+(3)}$	$(n-2\,0\,0\,0)$	$n \geqslant 2$	$5 + n/2$

[‡1] This notation is explained in ref. [16].

are two states for which n may take the value -1 ("the basement"). These two states constitute the $N = 8$ singleton multiplet [14,17,18] but the corresponding modes on the round seven-sphere vanish identically. Therefore they certainly do not appear at the linearised level. In what follows we will also need the formulas for the dimension and the Casimir of an irreducible SO(8) representation (a, b, c, d). They are given by [12]

$$\begin{aligned} D(a,b,c,d) = {} & (1+a)(1+b)(1+c)(1+d)\left(1+\tfrac{1}{2}(a+b)\right)\left(1+\tfrac{1}{2}(b+c)\right)\left(1+\tfrac{1}{2}(b+d)\right) \\ & \times\left(1+\tfrac{1}{3}(a+b+c)\right)\left(1+\tfrac{1}{3}(a+b+d)\right)\left(1+\tfrac{1}{3}(b+c+d)\right) \\ & \times\left(1+\tfrac{1}{4}(a+b+c+d)\right)\left(1+\tfrac{1}{5}(a+2b+c+d)\right), \end{aligned} \tag{2}$$

and

$$\begin{aligned} C(a,b,c,d) = {} & \tfrac{1}{28} D(a,b,c,d)\left[a\left(a+b+\tfrac{1}{2}(c+d)+6\right)+b(a+2b+c+d+10)\right. \\ & \left.+c\left(\tfrac{1}{2}(a+d)+b+c+6\right)+d\left(\tfrac{1}{2}(a+c)+b+d+6\right)\right]. \end{aligned} \tag{3}$$

It is now straightforward to determine the contribution of the massive states to the one-loop β-function using the above results. For arbitrary n. One simply inserts (3) into (1) and sums all the contributions in a given Osp(8|4) multiplet, i.e. for fixed n. We emphasize that for massive states the associated Higgs-states of lower helicity must be taken into account. As a consequence of (2), (3), this expression is an eighth-order polynomial in the "floor"-label n, and we have therefore resorted to the algebraic programme REDUCE for its actual evaluation. The result is that the contributions cancel within each Osp(8|4) multiplet, i.e. "floor-by-floor".

An alternative method of performing this calculation is by ζ-function regularization. This method is more elaborate in that it requires the knowledge of the energy spectrum of a given SO(3.2) representation [19] but it has the advantage that it allows the determination of the Casimir energies as well. In figs. 1 and 2 we have reproduced the weight diagrams of the unitary irreducible massless representations of SO(3.2). These are labelled by the energy E_0 of the state of lowest energy and by its total spin s. Each dot in the weight diagram represents $2j+1$ states where $j = l + s$ denotes the total angular momentum of the corresponding states. For a massless representation, the representation labels. E_0 and s obey the relation $E_0 = s + 1$. Note that the weight diagram for $s = 0$ differs from the weight diagrams for higher spin in that it contains only "half as many" states. The weight diagrams for massive spin $s > 1/2$, for which $E_0 > s + 1$, are obtained by adjoining the lower helicity Higgs states such that the lowest energy is increased by one

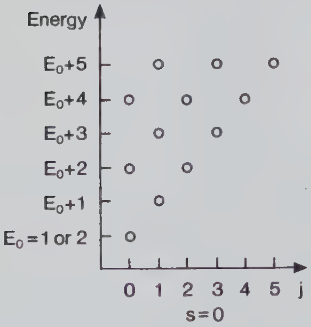


Fig. 1. The massless spin-zero representations. The diagram is the same for massive spin-zero representations for which $E_0 \neq 1, 2$.

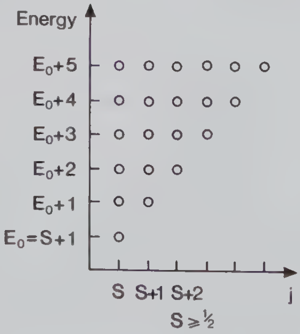


Fig. 2. The massless representations for spin greater than 0.

unit as the spin is decreased by one unit. For example, the weight diagram for a massive spin-one particle is the superposition of helicity-one and helicity-zero states shown in fig. 3. The zero-point energy corresponding to these weight diagrams is given by a formally divergent sum. In order to regularize it, we define generalized ζ-functions as follows. For spin $s > 0$, we define

$$\begin{aligned}\zeta(z; E_0, s) &= \sum_{k,l=0}^{\infty} [2(l+s)+1](E_0+k+l)^{-z}\\ &= \zeta(z-2, E_0) + 2(s+1-E_0)\zeta(z-1, E_0) + (E_0-1)(E_0-2s-1)\zeta(z, E_0),\end{aligned} \tag{4}$$

where

$$\zeta(z, \alpha) = \sum_{m=0}^{\infty} (m+\alpha)^{-z}, \quad \operatorname{Re} z > 1 \tag{5}$$

is the usual generalized ζ-function (see e.g. ref. [20]). Formula (4) was first obtained by Hull [21]. From (4) and (5), one immediately verifies that $\zeta(z; E_0, s)$ is holomorphic in the region $\operatorname{Re} z > 3$ of the complex plane. To the left of this region, the ζ-function is defined by analytic continuation. For the case of spin-zero, the correct definition reads

$$\begin{aligned}\tilde{\zeta}(z; E_0) &= \sum_{k,l=0}^{\infty} (2l+1)(E_0+2k+1)^{-z}\\ &= \tfrac{1}{2}\left[\zeta(z-2, E_0) + (3-2E_0)\zeta(z-1, E_0) + \left(E_0^2 - 3E_0 + 2\right)\zeta(z, E_0)\right].\end{aligned} \tag{6}$$

The range of summation in (4) and (6) is obtained by inspection of the corresponding weight diagrams in figs. 1 and 2. For the calculation of the Casimir energies, we have to evaluate these ζ-functions at $z = -1$. The relevant values of the ordinary ζ-function are given by [20]

$$\zeta(-n, \alpha) = -B_{n+1}(\alpha)/(n+1) \tag{7}$$

with

$$B_2(\alpha) = \alpha^2 - \alpha + \tfrac{1}{6}, \quad B_3(\alpha) = \alpha^3 - \tfrac{3}{2}\alpha^2 + \tfrac{1}{2}\alpha, \quad B_4(\alpha) = \alpha^4 - 2\alpha^3 + \alpha^2 - \tfrac{1}{30}. \tag{8}$$

Substituting these formulas into eqs. (4) and (6), we get for spin $s > 0$

$$\zeta(-1; E_0, s) = -\tfrac{1}{12}E_0^4 + \tfrac{1}{3}(s+1)E_0^3 - \tfrac{1}{2}(2s+1)E_0^2 + \tfrac{1}{6}(5s+2)E_0 - \tfrac{1}{120}(20s+9), \tag{9}$$

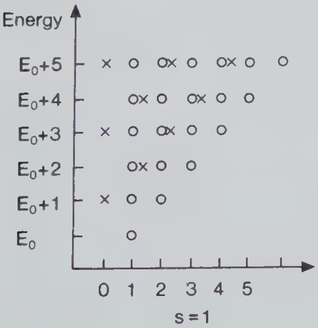


Fig. 3. Massive spin-one representations. The states marked by a cross × decouple for $E_0 = 2$.

which reduces to the correct result for $E_0 = s + 1$ [10], namely

$$\zeta(-1; s+1, s) = \tfrac{1}{4}s^4 - \tfrac{1}{6}s^2 + \tfrac{1}{120}. \tag{10}$$

For spin-zero particles on the other hand, we obtain

$$\tilde{\zeta}(-1; E_0) = -\tfrac{1}{24}E_0^4 + \tfrac{1}{4}E_0^3 - \tfrac{1}{2}E_0^2 + \tfrac{3}{8}E_0 - \tfrac{19}{240}. \tag{11}$$

These formulas are sufficient for the massless case. For the massive states, the contributions of the lower helicity states must be added. For massive spin-two this procedure yields the result

$$\zeta(z; E_0, 2) + \zeta(z; E_0 + 1, 1) + \tilde{\zeta}(z; E_0 + 2) \tag{12}$$

whereas, for massive spin $\frac{3}{2}$, it leads to

$$\zeta(z; E_0, \tfrac{3}{2}) + \zeta(z; E_0 + 1, \tfrac{1}{2}). \tag{13}$$

Collecting all these terms for a fixed Osp(8|4) multiplet characterized by the integer n, we arrive at the following expression

$$\begin{aligned}
\zeta_{\text{multiplet}}(z) = {} & D(n,0,0,0)\left[\zeta(z; 3+\tfrac{1}{2}n, 2) + \zeta(z; 4+\tfrac{1}{2}n, 1) + \tilde{\zeta}(z; 5+\tfrac{1}{2}n)\right] \\
& - D(n,0,0,1)\left[\zeta(z; \tfrac{5}{2}+\tfrac{1}{2}n, \tfrac{3}{2}) + \zeta(z; \tfrac{7}{2}+\tfrac{1}{2}n, \tfrac{1}{2})\right] - D(n-1,0,1,0)\left[\zeta(z; \tfrac{7}{2}+\tfrac{1}{2}n, \tfrac{3}{2}) + \zeta(z; \tfrac{9}{2}+\tfrac{1}{2}n, \tfrac{1}{2})\right] \\
& + D(n,1,0,0)\left[\zeta(z; 2+\tfrac{1}{2}n, 1) + \tilde{\zeta}(z; 3+\tfrac{1}{2}n)\right] + D(n-1,0,1,1)\left[\zeta(z; 3+\tfrac{1}{2}n, 1) + \tilde{\zeta}(z; 4+\tfrac{1}{2}n)\right] \\
& + D(n-2,1,0,1,1)\left[\zeta(z; 4+\tfrac{1}{2}n, 1) + \tilde{\zeta}(z; 5+\tfrac{1}{2}n)\right] - D(n+1,0,1,0)\zeta(z; \tfrac{3}{2}+\tfrac{1}{2}n, \tfrac{1}{2}) \\
& - D(n-1,1,1,0)\zeta(z; \tfrac{5}{2}+\tfrac{1}{2}n, \tfrac{1}{2}) - D(n-2,1,0,1)\zeta(z; \tfrac{7}{2}+\tfrac{1}{2}n, \tfrac{1}{2}) - D(n-2,0,0,1)\zeta(z; \tfrac{9}{2}+\tfrac{1}{2}n, \tfrac{1}{2}) \\
& + D(n+2,0,0,0)\tilde{\zeta}(z; 1+\tfrac{1}{2}n) + D(n,0,2,0)\tilde{\zeta}(z; 2+\tfrac{1}{2}n) + D(n-2,2,0,0)\tilde{\zeta}(z; 3+\tfrac{1}{2}n) \\
& + D(n-2,0,0,2)\tilde{\zeta}(z; 4+\tfrac{1}{2}n) + D(n-2,0,0,0)\tilde{\zeta}(z; 5+\tfrac{1}{2}n).
\end{aligned} \tag{14}$$

Again, using REDUCE we have shown that $\zeta_{\text{multiplet}}(z)$ vanishes at $z = -1$ for arbitrary n. In a similar way one may analyse the pole structure of $\zeta_{\text{multiplet}}(z)$. The generalized Reimann ζ-function has a single first order pole at $z = 1$, with residue one. Therefore the function $\zeta_{\text{multiplet}}(z)$ could, a priori, have poles at $z = 3$, 2, 1. By use of (4) and (6) together with (14) the residues of $\zeta_{\text{multiplet}}(z)$ at these values of z may be easily evaluated. Again they are found to vanish separately for each n. Hence $\zeta_{\text{multiplet}}(z)$ is an entire function for arbitrary n. This is significant because the residues depend upon the asymptotic density of states at high energy which enter the divergent integrals of the quantized theory.

As we mentioned above the states belonging to the $N = 8$ singleton multiplet do not appear at the linearized level. Nonetheless, one may enquire whether they contribute to the zero-point energy. In fact, their possible contribution to the Casimir energy may be calculated using the ζ-function method, whereas Curtright's method is inapplicable since these multiplets have no Poincaré limit. From the weight diagram for the singleton multiplet depicted in fig. 4 one may derive their contribution to the ζ-function as

$$\zeta_{\text{singleton}}(z) = 16\zeta(z-1, \tfrac{1}{2}) - 16\zeta(z-1, 1), \tag{15}$$

where a factor of 8 is included because the states belong to eight-dimensional representations of SO(8). Evidently the residues at $z = 2$ cancel and one easily checks from formulas (7) and (8) that $\zeta_{\text{singleton}}(-1) = 0$.

The vanishing of the β-function is certainly related to the absence of one-loop counter-terms in the eleven-dimensional theory [22]. It should be pointed out, however, that the Freund–Rubin parameter is not a parameter of the original lagrangian of the eleven-dimensional theory. The obvious analogy is that of quantum electrodynamics with an external electromagnetic field ("electrovac"). In order to render this

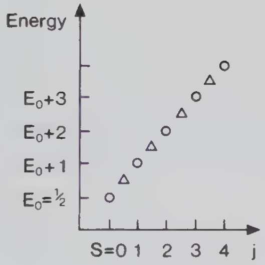


Fig. 4. Singleton spectrum: bosonic states are marked by a circle ○ and fermionic ones by a triangle △

theory finite at one loop, the background field must be renormalized in the same way as the fluctuations and therefore the renormalization properties of the theory in an electromagnetic background coincide with those in a vanishing background [23]. Although this argument indicates that the total contribution of the massless and the massive modes to the β-function vanishes, it is surprising to discover that the cancellations actually take place *within* each Osp(8|4) multiplet, i.e. "floor-by-floor". The argument also implies that the β-function will vanish at one loop for any other background. In some cases, this may be explicitly checked: for instance, using Curtright's formula, we have verified the vanishing at one loop for some of the stationary points of the $N=8$ potential.

While the one-loop contribution to the β-function does not depend upon the background around which one expands, the Casimir energies do depend upon it. Obviously it would be of interest to investigate this problem in a more general context but this requires knowledge of the full spectrum of fluctuations around compactified solutions other than the round seven-sphere which is not yet available. of special importance is the question of whether or not the classical relation between gauge coupling constant and cosmological constant is preserved in broken phases of the quantized theory. Our calculation provides the first example of a self-consistent solution of the one-loop corrected equations of motion of eleven-dimensional supergravity. In contrast to the suggestion in ref. [24], the resulting cosmological constant does not vanish in our case.

Finally we turn to the Euler number divergence. For the massless modes this is known not to vanish [7]. From this it is obvious that there can be no floor-by-floor cancellation. In fact one finds, using the results tabulated in ref. [7], that each floor contributes an amount equal to $D(n,0,0,0)$ times the contribution of the massless sector. The total contribution is therefore formally

$$-5\sum_{n=0}^{\infty} D(n,0,0,0). \tag{16}$$

Just as before this can be regularized by introducing yet another ζ-function by

$$\zeta_\Delta(z)=1+\sum_{n=1}^{\infty} D(n,0,0,0)[n(n+6)]^{-z}. \tag{17}$$

This ζ-function is that associated with the scalar laplacian Δ on the round seven-sphere whose eigenvalues are $n(n+6)$ with multiplicity $D(n,0,0,0)$. We are interested in $\zeta_\Delta(0)$ which represents the "renormalized" total number of modes and which is obtained by analytically continuing $\zeta_\Delta(z)$ from the region $\mathrm{Re}\, z>\frac{7}{2}$ where the series in (17) converges. By use of the following integral representation

$$\zeta_\Delta(z)=\frac{1}{\Gamma(z)}\int_0^\infty t^{z-1}\,\mathrm{Tr}\,\det(-t\Delta)\mathrm{d}t \tag{18}$$

it follows that $\zeta_\Delta(0)$ is proportional to the constant term in the asymptotic expansion of $\exp(-t\Delta)$ at small t. It is well known that this expansion is of the form $(4\pi t)^{-d/2}\times$ (series in positive integral powers of t)

where d is the dimension of the space over which Δ is defined [25]. Since in our case $d=7$ there is no constant term on the seven-sphere and thus [‡2]

$$\zeta_\Delta(0)=0. \tag{19}$$

Therefore the sum of the contributions of all modes to the Euler counterterm vanishes. That is, the massive modes exactly cancel the contribution of the massless sector. Similar arguments apply to any other seven-dimensional internal manifold

H.N. is grateful to DAMTP for the hospitality extended to him during his stay. G.W.G. would like to thank C.M. Hull for some useful discussions.

Note added in proof. After this letter had been accepted for publication, we learnt that similar results had been obtained independently by Inami and Yamagishi [26].

[‡2] A similar cancellation has been shown to take place in the simpler five-dimensional Kaluza–Klein theory [22].

References

[1] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. 76B (1978) 409.
[2] L. Castellani, R. D'Auria and P. Fré, in: Supersymmetry and supergravity 83, Proc. 19th Winter School on Theoretical physics (Karpacz) ed. B. Milewski (World Scientific, Singapore);
M.J. Duff, B.E.W. Nilsson and C.N. Pope, preprint ICTP/82-83/29 (1983);
B. de Wit, preprint NIKHEF-H/83-18 (1983);
P. Van Nieuwenhuizen, Les Houches Lecture Notes, to be published (1983).
[3] F. Englert and H. Nicolai, CERN preprint TH.3711 (1983).
[4] P.G.O. Freund and M.A. Rubin, Phys. Lett. 97B (1980) 233.
[5] M.J. Duff and C.N. Pope, in: Supergravity '82, eds. S. Ferrara, J.G. Taylor and P. Van Nieuwenhuizen (World Scientific, Singapore, 1983).
[6] B. Biran, B. de Wit, F. Englert and H. Nicolai, Phys. Singapore, Lett. 124B (1983) 45.
[7] S.M. Christensen, M.J. Duff, G.W. Gibbons and M. Roček, Phys. Rev. Lett. 45 (1980) 161
[8] T. Curtright, Phys. Lett. 102B (1981) 17.
[9] T. B. Allen and S. Davis, Phys. Lett. 124B (1983) 353.
[10] G.W. Gibbons, in: Differential geometric methods in mathematical physics II, eds. K. Bleuler, H.R. Petry and A. Reetz, Lecture Notes in Mathematics, Vol. 676 (Springer, Berlin).
[11] J.S. Dowker, Phys. Rev. 28D (1983) 3031.
[12] D.Z. Freedman and H. Nicolai, Nucl. Phys. B237 (1984) 342.
[13] B. Biran, A. Casher, F. Englert, M. Rooman and P. Spindel, Phys. Lett. 134B (1984) 179;
A. Casher, F. Englert, H. Nicolai and M. Rooman, CERN preprint TH. 3794 (1984).
[14] E. Sezgin, Phys. Lett. 138B (1984) 57.
[15] R. D'Auria and P. Fré, Torino preprint IFTT 430 (1983).
[16] R. Slansky, Phys. Rep. C 79 (1981) 1.
[17] M. Günaydin, in: Group theoretical methods in physics, Lecture Notes in Physics, Vol. 180 (Springer, Berlin, 1983).
[18] H. Nicolai and E. Sezgin, CERN preprint TH. 3845 (1984).
[19] C. Fronsdal, Rev. Mod. Phys. 37 (1965) 221; Phys. Rev. D10 (1974) 589; Phys. Rev. D12 (1975) 3819;
C. Fronsdal and R.B. Haugen, Phys. Rev. D12 (1975) 3819.
[20] W. Magnus, F. Oberhettinger and R.P. Soni, Formulae and theorems for the special functions of mathematical physics" (Springer, Berlin).
[21] C.M. Hull, DAMTP preprint (1983), to be published in Nucl. Phys. B.
[22] M.J. Duff and D. Toms, in: Unification of the fundamental particle interaction II, eds. J. Ellis and S. Ferrara (Plenum, New York, 1982).
[23] J. Schwinger, Phys. Rev. 82 (1951) 664.
[24] E.S. Fradkin and A.A. Tseytlin, Nucl. Phys. B227 (1983) 252;
P. Candelas and D.J. Raine, Texas preprint (1983).
[25] P. Gilkey, The index theorem and the heat kernel (Publish or Perish, Wilmington, DE, 1974).
[26] T. Inami and K. Yamagishi, Phys. Lett. 143B (1984) 115.

VANISHING QUANTUM VACUUM ENERGY IN ELEVEN-DIMENSIONAL SUPERGRAVITY ON THE ROUND SEVEN-SPHERE

Takeo INAMI
Research Institute for Fundamental Physics, Kyoto University, Sakyo-ku, Kyoto 606, Japan

and

Kengo YAMAGISHI [1]
Institute of Physics, University of Tokyo-Komaba, Meguro-ku, Tokyo 153, Japan

Received 8 May 1984

Quantum corrections to the vacuum energy are evaluated at one-loop order in eleven-dimensional supergravity on the round seven-sphere S^7 and are shown to vanish. The cancellation is also shown for all ultraviolet poles at $z = 11/2$, $10/2, \ldots,$ corresponding to divergences of eleventh and lower powers of momentum cut-off Λ.

The recent enthusiasm for extended supergravity is based on the hope that it may provide us a (perturbatively) finite theory of gravity. Ultraviolet finiteness has so far been proven for massless theories, at one- and two-loop orders in simple supergravity [1] and at one-loop order in gauged $O(N)$ (de Sitter) supergravity for $5 \leqslant N \leqslant 8$ (on shell) [2–5]. It has been shown in certain cases that remarkable cancellations occur for quartic and quadratic as well as logarithmic divergences by virtue of supersymmetry [6].

Gauge $O(N)$ supergravities are considered to be obtained as the massless sector of eleven-dimensional supergravity [7] by compactifying spontaneously the seven space-like dimensions [8–10]. In this letter we address ourselves to the question of how far the cancellation of ultraviolet infinities due to supersymmetry in the spontaneously compactified eleven-dimensional supergravity goes. As the simplest example we consider the eleven-dimensional supergravity compactified on the round seven-sphere without torsion [8,9]. We look at the ultraviolet behaviour of the vacuum energy at one-loop order. The round S^7 solution contains in addition to the massless sector a tower of an infinite number of sets of massive fields. These massive fields are given no place in low-energy effective theory, but one may envisage that the ultraviolet behaviour will be affected by the inclusion of such fields [11].

The total vacuum energy at one-loop order is obtained by summing zero-point energies.

$$E_{\text{vac}} = \tfrac{1}{2}\left(\sum E_{\text{boson}} - \sum E_{\text{fermion}}\right). \qquad (1)$$

Here E_{boson} and E_{fermion} are eigenvalues (at tree level) of the energy operator M_{04} in anti-de Sitter (AdS) space–time, and are not equal to each other in the Sitter supersymmetry. Hence the cancellation between the bosonic and fermionic contributions does not occur trivially, in contrast to the case of Poincaré supersymmetry. Rather, the summation in (1) appears strongly ultraviolet divergent (*to eleventh power* of momentum cut-off Λ [‡1]), and has to be regularized.

---

[1] Postdoctoral Fellow, Japan Society for the Promotion of Science.

[‡1] The summation over the (massive) harmonic modes gives rise to this type of divergences, although the theory is defined in four dimensions.

We use the zeta function method [12] and define a function $E_{\mathrm{vac}}(z)$ of a complex variable z [4],

$$E_{\mathrm{vac}}(z) \equiv \tfrac{1}{2}\left(\sum (E_{\mathrm{boson}})^{1-2z} - \sum (E_{\mathrm{fermion}})^{1-2z}\right). \tag{2}$$

The summation now converges for sufficiently large positive Re z, i.e., Re $z > 11/2$ in the present case, and defines an analytic function, regular except for possible poles at positive integer and half-integer values of z. Integer poles (half-integer poles) can be shown to represent ultraviolet divergences of even (odd) powers of Λ. The vacuum energy is given by analytically continuing $E_{\mathrm{vac}}(z)$ to $z = 0$, $E_{\mathrm{vac}} = E_{\mathrm{vac}}\ (z = 0)$. It gives the coefficient of logarithmic divergence in the effective action.

To perform the summation in (2) one has to know the energy eigenvalues. Here we give a summary of the complete energy spectrum of the eleven-dimensional supergravity on the round S^7. The maximal invariance of the round S^7 solution without torsion is $N = 8$ de Sitter supersymmetry, OSp(8,4) [13]. The energy spectrum can be classified into irreducible representations of this algebra.

In AdS space–time a particle corresponds to an infinite dimensional unitary representation $D(E_0, s)$ [14], where E_0 is the lowest value of energy and s is the angular momentum of this state, which is referred to as the particle spin. The energy within this multiplet is quantized and is given by (in the unit of inverse radius of the sphere)

$$E_{kjh} = E_0 + k + j - s + 2(s - |h|), \tag{3}$$

where k, j and h are radial quantum number, total angular momentum and helicity ($= s_z$), respectively,

$$k = 0, 1, 2, \ldots \quad (0, 2, 4, \ldots \text{ for } h = 0), \quad j = |h|, |h| + 1, \ldots, \quad h = -s, -s + 1, \ldots, s. \tag{4}$$

Note the additional $2j + 1$ degrees of degeneracy due to the z-component of the total angular momentum.

Eleven-dimensional supergravity on the round S^7 contains an infinite tower of sets of AdS multiplets of spin $s = 2, 3/2, \ldots, 0$. The lowest values of energy of these multiplets are denoted by $E_0^{(n)} \equiv \omega_n$ for the nth mode of the S^7 harmonic series. They can be found by diagonalizing mass operators appearing in the linearized theory [15,16]. The resulting particle content is given in terms of the AdS and SO(8) representations for the nth massive/massless mode as follows,

$$\begin{aligned}
&D(n/2 + 3, 2) \times (n\ 0\ 0\ 0) \oplus D(n/2 + 5/2, 3/2) \times (n\ 0\ 0\ 1) \oplus D(n/2 + 7/2, 3/2) \times (n - 1\ 0\ 1\ 0)\\
&\oplus D(n/2 + 2, 1) \times (n\ 1\ 0\ 0) \oplus D(n/2 + 3, 1) \times (n - 1\ 0\ 1\ 1) \oplus D(n/2 + 4, 1) \times (n - 2\ 1\ 0\ 0)\\
&\oplus D(n/2 + 3/2, 1/2) \times (n + 1\ 0\ 1\ 0) \oplus D(n/2 + 5/2, 1/2) \times (n - 1\ 1\ 1\ 0) \oplus D(n/2 + 7/2, 1/2)\\
&\quad \times (n - 2\ 1\ 0\ 1)\\
&\oplus D(n/2 + 9/2, 1/2) \times (n - 2\ 0\ 0\ 1) \oplus D(n/2 + 1, 0) \times (n + 2\ 0\ 0\ 0) \oplus D(n/2 + 2, 0) \times (n\ 0\ 2\ 0)\\
&\oplus D(n/2 + 3, 0) \times (n - 2\ 2\ 0\ 0) \oplus D(n/2 + 4, 0) \times (n - 2\ 0\ 0\ 2) \oplus D(n/2 + 5, 0) \times (n - 2\ 0\ 0\ 0).
\end{aligned} \tag{5}$$

Here $(a\ b\ c\ d)$ is the Dynkin label of the gauge group SO(8). (States with a label including negative integers should be discarded.) Note that the ground state ($n = 0$) mode is composed of massless AdS multiplets $(\omega_0 + s + 1)$ with pure helicity states $h = \pm s$.

For later use we classify the massive states in (5) in terms of helicity states. To this end we use a slightly modified definition of lowest energy,

$$\tilde{\omega}_n \equiv \omega_n - s + 2(s - |h|). \tag{6}$$

The massless mode ($n = 0$) has the common value $\tilde{\omega}_0 = 1$. The $\tilde{\omega}_n$ of massive modes ($n \neq 0$) can be written as

$$\tilde{\omega}_n = n/2 + \delta, \quad \delta = 1, \ldots, 5. \tag{7}$$

The SO(8) Dynkin label for each of the helicity states in $D(\omega_n, h)$ for the nth harmonic mode are given in table 1.

The multiplicity $d_n(\delta, h)$ of the states in table 1 can be found from the formula:

$$\begin{aligned}\dim(a\,b\,c\,d) = {} & (1+a)(1+b)(1+c)(1+d)[1+(a+b)/2][1+(b+c)/2][1+(b+d)/2] \\ & \times [1+(a+b+c)/3][1+(a+b+d)/3][1+(b+c+d)/3] \\ & \times [1+(a+b+c+d)/4][1+(a+2b+c+d)/5]. \end{aligned} \tag{8}$$

The d_n turns out to be a sixth order polynomial in n and can be written as

$$d_n(\delta, h) = \sum_{i=0}^{6} a_i(\delta, h) n^i. \tag{9}$$

The $a_i(\delta, h)$ are given in table 2. To allow a supersymmetric classification of AdS multiplets, one has to combine two helicity 0 states $D(n/2+\delta, 0)$ and $D(n/2+\delta+1, 0)$ $[0 \leqslant \delta \leqslant 4]$ into one [14]. The resulting helicity 0 states can be denoted by $(\delta, 0)$ $[0 \leqslant \delta \leqslant 4]$ and follow the same energy spacing rule as other helicity states.

By examination of the table of a_i we can readily verify the following two kinds of sum rules involving $a_i(\delta, h)$ for fixed i.

$$\sum_{h=-2}^{2} (-)^{2h} a_i(\delta, h) = 0, \qquad \sum_{h=-2}^{2} (-)^{2h} h^k \left(\sum_{\delta=1}^{4} a_i(\delta, h) \right) = 0 \quad (k = 0, 1, \dots, 7). \tag{10,11}$$

Obviously, the same forms of sum rule hold for $d_n(\delta, h)$ for fixed n in place of $a_i(\delta, h)$. The first and second sum rules are spin sum rules of $N=1$ and $N=8$ supersymmetry, respectively, as pointed out by Curtright some time ago [3,17]. It is remarkable that the subset of states (δ, h) of fixed δ(those on a single row in table 1) form an $N=1$ "supermultiplet". (This nice property would not be manifest, if the modified definition $\tilde{\omega}_n$ of the lowest energy were not used.) The second sum rule (11) implies that the whole set of states (δ, h) in table 1 form an $N=8$ supermultiplet of OSp(8,4). To prove this conjecture one needs to deal with transformation properties of the fields corresponding to this set of massive states.

We are now ready to evaluate the function $E_{\mathrm{vac}}(z)$. It is given by an infinite sum of $E_{\mathrm{vac}}^{(n)}(z)$, the one-loop contribution from the nth mode in the tower of S^7 harmonic series.

$$E_{\mathrm{vac}}(z) = \sum_{n=0}^{\infty} E_{\mathrm{vac}}^{(n)}(z), \quad E_{\mathrm{vac}}^{(n)}(z) = \tfrac{1}{2} \sum_{h=-2}^{2} (-)^{2h} \sum_{\delta=1}^{4} d_n(\delta, h) \sum_{k=0}^{\infty} \sum_{j=|h|}^{\infty} (2j+1)(\tilde{\omega}_n + k + j)^{1-2z}. \tag{12}$$

This expression is strongly ultraviolet divergent. There are two possible sources of infinities. One is the infinite summation over k and j, the four-momentum summation, and the other is that over the harmonic modes n.

The first type of infinities can be extracted by rearranging the summation over k and j so as to give the generalized Riemann zeta function $\zeta(u, a)$.

$$\begin{aligned} E_{\mathrm{vac}}^{(n)}(z) = {} & \sum_{h} (-)^{2h} \sum_{\delta} d_n(\delta, h) \big\{ \zeta(2z-3, \tilde{\omega}_n + |h|) - 2(\tilde{\omega}_n - 1)\zeta(2z-2, \tilde{\omega}_n + |h|) \\ & + \left[(\tilde{\omega}_n - 1)^2 - h^2 \right] \zeta(2z-1, \tilde{\omega}_n + |h|) \big\} / 2 \end{aligned} \tag{13}$$

$\zeta(u, a)$ is a meromorphic function of u and has a pole at $u=1$, $\lim_{u \to 1} \zeta(u, a) = (u-1)^{-1}$. The three terms in the curly brackets in (12) have poles at $z = 2$, $3/2$ and 1, respectively, corresponding to quartic, cubic

Table 1
Helicity states in the nth OSp(8,4) multiplet.

$\tilde{\omega}_n$	h				
	0	1/2	1	3/2	2
$n/2+1$	$(n+2\,0\,0\,0)$	$(n+1\,0\,1\,0)$	$(n\,1\,0\,0)$	$(n\,0\,0\,1)$	$(n\,0\,0\,0)$
$n/2+2$	$(n\,0\,2\,0)$	$(n-1\,1\,1\,0)$	$(n-1\,0\,1\,1)$	$(n-1\,0\,1\,0)$	
$n/2+3$	$(n\,1\,0\,0)$ $+(n-2\,2\,0\,0)$	$(n-2\,1\,0\,1)$ $+(n\,0\,0\,1)$	$(n-2\,1\,0\,0)$ $+(n\,0\,0\,0)$		
$n/2+4$	$(n-2\,0\,0\,2)$ $+(n-1\,0\,1\,1)$	$(n-2\,0\,0\,1)$ $+(n-1\,0\,1\,0)$			
$n/2+5$	$(n-2\,0\,0\,0)$ $+(n-2\,1\,0\,0)$ $+(n\,0\,0\,0)$				

Table 2

δ	h					
	0	1/2	1	3/2	2	
1	1	4	6	4	1	
2	9	20	15	4		
3	17	24	7			
4	8	8				$a_6 \times 3\cdot 5!$
1	5	18	24	14	3	
2	35	70	45	10		
3	49	64	15			
4	16	16				$a_5 \times 3\cdot 5!/6$
1	37	118	138	70	13	
2	189	338	183	34		
3	185	228	43			
4	44	44				$a_4 \times 3\cdot 5!/10$
1	40	111	112	49	8	
2	136	217	96	15		
3	88	104	16			
4	16	16				$a_3 \times 3\cdot 5!/60$
1	8629	20416	17334	6496	949	
2	16821	23960	8235	1096		
3	6773	7776	1003			
4	992	992				$a_2 \times 3\cdot 5!$
1	2715	5352	3656	1176	157	
2	2205	2800	675	80		
3	511	576	65			
4	64	64				$a_1 \times 3\cdot 5!/6$
1	35	56	28	8	1	
2–4	0	0				$a_0 \times 3\cdot 5!/360$

and quadratic divergences. The pole part is

$$E_{\text{vac}}^{(n)}(z)=\tfrac{1}{2}\sum_{h}(-)^{2h}\sum_{\delta}d_n(\delta,h)\left\{\tfrac{1}{2}(z-2)^{-1}-(\tilde{\omega}_n-1)(z-3/2)^{-1}+\tfrac{1}{2}\left[(\tilde{\omega}_n-1)^2-h^2\right](z-1)^{-1}\right\}. \quad (14)$$

It is easy to see that the residues of the three pole terms vanish thanks to the spin sum rule (10) and (11). We have demonstrated the cancellation of ultraviolet divergences of fourth and lower powers. Remarkably, this cancellation takes place mode by mode in the harmonic series.

We next consider the second type of infinity. The S^7 mode summation in (12) potentially gives rise to poles of $E_{\text{vac}}(z)$ at $z \leqslant 11/2$, corresponding to ultraviolet divergences of eleventh and lower powers. These pole terms can be extracted by substituting the asymptotic expansion of $\zeta(u,\omega_n+|h|)$ for large n [18]. We have been able to show by explicit computations using table 2 that the all residues of the pole terms vanish. The residues of the first three poles at $z=11/2$, 5, 9/2 vanish thanks to the spin sum rules (10) and (11). As we go to lower poles, the spin sum rules are not sufficient for the cancellation of their residues. We have found an additional set of sum rules for $a_i(\delta,h)$ involving moments of both spin h and energy δ, which are responsible for vanishing residues of poles at $z \leqslant 4$.

$$\sum_{h,\delta}(-)^{2h}(4|h|^3-|h|+6\delta h^2)a_i(\delta,h)=0 \quad (2\leqslant i\leqslant 6), \quad (15)$$

$$\sum_{h,\delta}(-)^{2h}(\delta(4|h|^3-|h|)+3\delta^2h^2)a_i(\delta,h)=0 \quad (3\leqslant i\leqslant 6), \quad (16)$$

$$\sum_{h,\delta}(-)^{2h}\left[4|h|^5-|h|^3+\delta(15h^4-4h^2)+5\delta^2(4|h|^3-|h|)+10\delta^3h^2\right]a_i(\delta,h)=0 \quad (4\leqslant \mathrm{i}\leqslant 6), \quad (17)$$

$$\sum_{h,\delta}(-)^{2h}\left[6\delta(4|h|^5-|h|^3)+3\delta^2(15h^4-4h^2)+10\delta^3(4|h|^3-|h|)+15\delta^4h^2\right]a_i(\delta,h)=0 \quad (\mathrm{i}=5,6). \quad (18)$$

The meanings of these new types of sum rules are yet unclear.

We finally proceed to the evaluation of the finite part of E_{vac}. Noting $\zeta(-l,a)=-B_{l+1}(a)/(l+1)$ [$l=0$, 1, 2,...], where $B_l(a)$ is Bernoulli's polynomial, we obtain from (13)

$$\begin{aligned}E_{\text{vac}}^{(n)}(0)=-\tfrac{1}{2}\sum_{h}(-)^{2h}\sum_{\delta}d_n(\delta,h)\Big[&B_4(\tilde{\omega}_n+|h|)/4-2(\tilde{\omega}_n-1)B_3(\tilde{\omega}_n+|h|)/3\\&+\left[(\tilde{\omega}_n-1)^2-h^2\right]B_2(\tilde{\omega}_n+|h|)/2\Big].\end{aligned} \quad (19)$$

After discarding terms which vanish after use of the spin sum rules (10) and (11), we have

$$E_{\text{vac}}^{(n)}(0)=-\tfrac{1}{2}\sum_{h}(-)^{2h}\sum_{\delta}d_n(\delta,h)\left[(\tilde{\omega}_n-1)(4|h|^3-|h|)/6+(\tilde{\omega}_n-1)^2h^2/2\right]. \quad (20)$$

Curiously enough, the contributions of $a_i(\delta,h)$ $(3\leqslant i\leqslant 6)$ turn out to vanish by virtue of the spin-energy sum rules (15) and (16). The contributions of the remaining $a_i(\delta,h)$ can be seen to cancel out among themselves. This complete our proof of vanishing vacuum energy at one-loop order.

The absence of ultraviolet divergences (at least) at one-loop order in the eleven-dimensional supergravity on the round S^7 solution, which we have demonstrated in this letter, is a consequence of the $N=8$ supersymmetry of this solution. It will be of interest to study whether the same good ultraviolet property is shared by other kinds of S^7 solutions with a fewer-N supersymmetry [10].

In the course of writing this letter, we have received a preprint by Gibbons and Nicolai [19], in which they have obtained virtually the same result as ours. We have shown in addition the cancellation of the ultraviolet poles at $z=11/2$, $10/2,\ldots$, $5/2$.

Note added. The vanishing Casimir energy in the eleven-dimensional supergravity on the round S^7 was conjectured previously by Duff [20]. He anticipated that the spin sum rule derivation of the vanishing β-function properties in $N > 4$ gauged supergravity might carry over when the massive OSp(8,4) multiplets are included.

References

[1] M.T. Grisaru, P. van Nieuwenhuizen and J.A. Vermaseren, Phys. Rev. Lett. 37 (1976) 1662;
M.T. Grisaru and H. Pendleton, Nucl. Phys. B124 (1977) 81.
[2] S.M. Christensen, M.J. Duff, G.W. Gibbons and M. Roček, Phys. Rev. Lett. 45 (1980) 161.
[3] T. Curtright, Phys. Lett. 102B (1981) 17.
[4] B. Allen and S. Davis, Phys. Lett. 124B (1983) 353.
[5] M.T. Grisaru and W. Siegel, Nucl. Phys. B201 (1982) 474;
K.S. Stelle and P.K. Townsend, Phys. Lett. 113B (1982) 25;
E.S. Fradkin and A.A. Tseytlin, Phys. Lett. 117B (1982) 303.
[6] L. Castellani and P. van Nieuwenhuizen, Nucl. Phys. B213 (1982) 305.
[7] E. Cremmer, B. Julia and B. Scherk, Phys. Lett. 76B (1978) 409.
[8] P.G.O. Fruend and M.A. Rubin, Phys. Lett. 97B (1980) 233.
[9] M.J. Duff and C.N. Pope, in: Supergravity '82 eds. S. Ferrara, J.G. Taylor and P. van Nieuwenhuizen (World Scientific, Singapore, 1983).
[10] F. Englert, Phys. Lett. 119B (1982) 339;
M.A. Awada, M.J. Duff and C.N. Pope, Phys. Rev. Lett. 50 (1983) 294;
L. Castellani, R. D'Auria and P. Fré, Torino preprint IFTT 427 (1983).
[11] M.J. Duff and D.J. Toms, CERN preprints TH.3259, 3248 (1982).
[12] S.W. Hawking, Commun. Math. Phys. 55 (1977) 33.
[13] R. D'Auria and P. Fré, Phys. Lett. 121B (1983) 141.
[14] C. Fronsdal, Phys. Rev. D12 (1975) 3819;
J. Fang and C. Fronsdal, Lett. Math. Phys. 2 (1978) 391;
S.J. Avis, C.J. Isham and D. Story, Phys. Rev. D18 (1978) 3565;
P. Breitenlohner and D.Z. Freedman, Ann. Phys. (NY) 144 (1982) 249.
[15] B. Biran, A. Casher, F. Englert, M. Rooman and P. Spindel, Phys. Lett. 134B (1984) 179;
A. Casher, F. Englert, H. Nicolai and M. Rooman, CERN preprint TH.3794 (1984);
E. Sezgin, Phys. Lett. 138B (1984) 57.
[16] W. Heidenreich, Phys. Lett. 110B (1982) 461;
D.Z. Freedman and H. Nicolai, Nucl. Phys. B237 (1984) 342.
[17] M.J. Duff, in: Supergravity '81 eds. S. Ferrara and J.G. Taylor (Cambridge U.P., London, 1982).
[18] A. Erdelyi et al., Higher transcendetal functions, Vol. 1 (McGraw-Hill, New York, 1953) p. 48.
[19] G.W. Gibbons and H. Nicolai, Phys. Lett 143B (1984) 108.
[20] M.J. Duff., Nucl. Phys. B219 (1983) 389.

CHAPTER

8

The Problem of Chiral Fermions

FERMION QUANTUM NUMBERS IN KALUZA-KLEIN THEORY[*]

Edward Witten
Joseph Henry Laboratories
Princeton University
Princeton, New Jersey 08544

ABSTRACT

The problem of obtaining left-right asymmetry of fermion quantum numbers in Kaluza-Klein theory is discussed. In the absence of elementary gauge fields, a theorem by Atiyah and Hirzebruch states that the Dirac equation in 4+n dimensions always leads to vector-like fermion quantum numbers in four dimensions. The proof of this theorem is sketched. It is shown that the same holds for the Rarita-Schwinger operator on homogeneous spaces but a general impossibility theorem for the Rarita-Schwinger field is not proved. (However, in view of the apparent restriction of supergravity to $d \leq 11$ this line of approach is severely constrained.) Also discussed are some Kaluza-Klein theories with elementary gauge fields, some difficulties in obtaining massless charged scalars, and some speculations about the cosmological constant.

---

[*] Research supported in part by the National Science Foundation under grant No. PHY80-19754.

I. INTRODUCTION

Kaluza-Klein theory[1] has recently attracted increasing interest as a program for unifying gauge interactions with gravity. This theory can be viewed[2] in terms of spontaneous symmetry breaking, the compact and non-compact dimensions being on an equal footing as far as the laws of nature are concerned, just as the photon and the massive vector mesons are treated symmetrically in the standard weak interaction models. This simple observation is one of the chief reasons for the revived interest[3] in Kaluza-Klein theory.

One may start with a general relativistic theory in 4+n dimensions, and assume the grand state to be $M^4 \times B$, where M^4 is four dimensional Minkowski space and B is a compact space. Continuous symmetries of B* will always be observed[4] as gauge symmetries in the effective four dimensional world. The gauge fields (which in general can be more numerous than the extra dimensions) originate in the normal mode expansion of the fluctuations in the 4+n dimensional metric tensor. For instance, starting in eleven or more dimensions, one can [4,5] obtain gauge fields of $SU(3) \times SU(2) \times U(1)$. Much of this paper will be devoted to the consequences of assuming that all observed gauge forces originate in this way, as part of the metric tensor, from a theory which originally had no elementary gauge fields. However, this is not necessarily the only attractive possibility. It might be equally attractive to start with a unified theory (perhaps a supergravity[6] or superstring[7] theory) that determines the original gauge group. (For instance, one of the n=2 supergravity theories in ten dimensions requires the existence of a U(1) gauge field.) Our remarks in section VI will be relevant to such theories.

---

*I.e., transformations that leave fixed both the geometry of B and the expectation values of any matter fields that may be present.

As soon as one begins to think about Kaluza-Klein theory, one faces a bewildering variety of choices. There are many assumptions one might make, and many facts about elementary particle physics one might try to explain. It appears unlikely that at the present time we can guess correctly the whole detailed form of the 4+n dimensional laws and all the key points of the dynamics. For these reasons, it seems important to isolate problems that can be addressed without claiming to understand all the details of a theory. As will become apparent, the problem of trying to predict the quark and lepton quantum numbers is such a problem, and it will be our main interest in this paper. However, we shall also make some remarks on certain other qualitative problems: the gauge hierarchy problem and the problem of the cosmological constant.

Since we will deal mainly with the problem of the fermion quantum numbers, it is worth while to briefly recall some aspects of that problem as it presently appears.

One of the most striking aspects of particle physics is that left handed fermions transform under S(3)×SU(2)×U(1) differently from the way the right handed fermions transform. (The quantum numbers are not "vector-like".)

For instance, left handed quarks are SU(2) doublets but right handed quarks are SU(2) singlets. Equivalently, one may say that the fermions of given helicity form a complex representation of SU(3)×SU(2)×U(1). The fermions of one generation transform under SU(3)×SU(2)×U(1) as $(3,2)^{1/3} \oplus (\bar{3},1)^{-4/3} \oplus (\bar{3},1)^{2/3} \oplus (1,1)^{2} \oplus (1,2)^{-1}$ which is a so-called complex representation (in other words, it is not equivalent to its complex conjugate, which is $(3,2)^{-1/3} \oplus (3,1)^{4/3} \oplus (3,1)^{-2/3} \oplus (1,1)^{-2} \oplus (1,2)^{1}$; by CPT, this is the representation furnished by the right-handed fermions). This fact is of utmost importance, because it means that bare masses of the quarks and leptons are forbidden by gauge invariance. The quarks and leptons can acquire mass only when SU(2)×U(1) is spontaneously broken. This, in turn, means that the quarks and leptons cannot be much heavier than the mass scale at which SU(2)×U(1) is broken; they cannot have masses of order, say, the Planck mass. The relative "lightness" of the fermions would therefore be explained if the "smallness" of the SU(2)×U(1) breaking scale were understood; it is not an independent problem. In this paper we will assume, in accord with the "survival hypothesis"[8] that the only light fermions are fermions that are required to be light by gauge invariance; this assumption will not always be explicitly stated.

The fact that the quantum numbers are not vector-like means that the spectrum of light fermions[*] depends only on the "universality class" of an SU(3)×SU(2)×U(1) invariant theory. The lightness of the light fermions and their quantum numbers cannot be modified by any SU(3)×SU(2)×U(1) invariant perturbations. We do not know at what length scale the spectrum of light fermions is determined, but it may be that this reflects physics at the smallest length scales.

Of course, there is no experimental proof that mirror fermions with V+A couplings to the usual W mesons will not be found, restoring the vector-like nature of the fermion spectrum. But there are many reasons to doubt that this will occur. If mirror fermions are discovered, we will lose our theoretical understanding of why the quarks and leptons are $\lesssim M_W$ in mass. (Of course, in any case we don't understand why some of the fermions are so much lighter than M_W.) If mirror fermions do exist they fail to a remarkable extent to mix with the usual fermions; the first generation fermions are very light and have almost pure V-A weak interactions. If mirrors do exist it is odd that none of the fourteen SU(3)×SU(2)×U(1) multiplets observed so far appear to be mirrors. This is all the more remarkable in that the mirrors cannot weigh more than at most a few hundred GeV (since they do not mix with usual fermions, their bare masses are forbidden by SU(2)×U(1)). Finally, and perhaps most convincingly, the triangle anomalies cancel among the observed quarks and leptons. This cancellation appears to be a rather striking confirmation of current ideas, but if mirrors exist it is just an elaborate and unnecessary charade, since the mirrors would automatically cancel the anomalies of the known fermions, whatever those anomalies might be.

---

[*] By the "spectrum of light fermions" we means the quantum numbers of the light fermions. We will usually speak of the quarks and leptons as if they were massless, ignoring the SU(2)×U(1) breaking.

What is more, the fermion representation is complicated (each family consists of five irreducible representations of SU(3)×SU(2)×U(1)) and redundant (there are three families). On the first point, no doubt the SU(5) and O(10) grand unified theories are the most successful efforts to date (a family is $\bar{5}_L+10_L$ in SU(5), or 16_L in O(10)). On the second point -- which is an updated version of Rabi's question "who ordered the muon?" -- there is no equally convincing answer. It is natural to try to embed the three families as one irreducible representation of a bigger group. Perhaps the most attractive such idea is to use the spinor representation[9] of O(N) for $N \geq 18$. The spinor representation of O(N) is the representation space of N gamma matrices, which automatically furnishes a representation of any subset of the gamma matrices; so the O(N) spinor transforms as a sum of spinor representations of any minimally embedded O(k) subgroup. For instance, the irreducible spinor of O(18) transforms under O(10) as four families plus (unfortunately) four antifamilies. This beautifully achieves the desired multiplicity, but it is not easy to eliminate the antifamilies. One may invoke a "hypercolor" force which becomes strong at ~1 Tev, breaking SU(2)×U(1) and confining the antifamilies. This elegant idea [10] has innumerable difficulties in detail. In this paper, difficulties will be much more conspicuous than phenomenological successes, but we will note, in section VI, that Kaluza-Klein theory gives an alternative way to avoid antifamilies in the O(N) approach to the family problem.

A discussion of zero modes of non-trivial Dirac operators in Kaluza-Klein theory was apparently first given in a special situation by Palla.[11] In (4), in connection with a discussion of some pseudo-realistic models, the detailed proposal was made that the quark and lepton quantum numbers are determined by the topology of a manifold with SU(3)×SU(2)×U(1) symmetry. The importance and difficulty of obtaining a complex representation were pointed out. Chapline and Slansky and Manton[12] discussed the problem of obtaining a complex spectrum in Kaluza-Klein theory; they anticipated the kinematical analysis of section II and some of the ideas of section VI. The kinematical analysis has been recently developed in much more detail by Wetterich,[13] who worked out all of the kinematical consequences of the mod 8 periodicity of the spinor represenation of O(N). He also introduced in a different language the mathematical concept of the character valued index, which as we will see plays a very important role. Models exhibiting many of the ideas of section VI have been analyzed by Randjbar-Daemi, Salam, and Strathdee.[14*] As regards

---

[*] I understand, in addition, that these authors have considered (unpublished) some of the detailed models in section VI.

fermion quantum numbers, the novelty in the present paper is primarily the restrictions disussed in sections IV and V and the more realistic models in section VI.

Our conclusions will be as follows. If all gauge fields are part of the metric tensor, then a theorem of Atiyah and Hirzebruch[15] states that the Dirac operator in 4+n dimensions always leads to vector-like quantum numbers in four dimensions. (The relevance of this theorem to Kaluza-Klein theory was first noted in reference (16).) For the Rarita-Schwinger operator the situation is more complicated. We will show that if the hidden dimensions form a homogeneous space the Rarita-Schwinger operator likewise always leads to vector-like quantum numbers. What happens in general for the Rarita-Schwinger operator on spaces that are not homogeneous I do not know. However, the fact that supergravity is apparently restrictd to $d \leq 11$,[17] and certain other facts discussed in section V, indicate that this avenue is not promising. If one is less ambitious and introduces elementary gauge fields in 4+n dimensions, it is possible but still subtle to get complex representations. Indeed, as we will see in section VI, one can naturally get very big, complicated, duplicated representations. In section VII, we discuss some other ways that the assumptions might be modified.

We will encounter considerable difficulties in our attempts to interpret the fermion quantum numbers as the solution of an index problem. Nevertheless, this seems to be a quite attractive idea.

II. PRELIMINARIES

Let us first recall how -- in a Kaluza-Klein theory with ground state $M^4 \times B$ -- massless particles originate as zero modes of appropriate wave operators on B. A massless Dirac particle in 4+n dimensions obeys

$$0 = \not{D}\psi = \sum_{i=1}^{4+n} \Gamma^{\mu} D_{\mu} \psi \tag{1}$$

where Γ^{μ}, $i=1 \ldots 4+n$, are the gamma matrices. Notice that we may as well use the minimal Dirac equation. Even if non-minimal terms (couplings to matter fields or non-minimal couplings to gravity) are present, they cannot change the quantum numbers of massless fermions in a complex representation of the symmetry group. This is an illustration of the fact that the problem of fermion quantum numbers depends only on the "universality class" of a theory. We do not have to believe we know which Dirac operator is physically relevant. If we define the four dimensional Dirac operator $\not{D}^{(4)} = \sum_{\mu=1}^{4} \Gamma^{\mu} D_{\mu}$ and the internal Dirac operator $\not{D}^{(n)} = \sum_{j=5}^{4+n} \Gamma^{j} D_{j}$, then (1) becomes

$$0 = \not{D}^{(4)} \psi + \not{D}^{(n)} \psi \tag{2}$$

We see that $\not{D}^{(n)}$ is the mass operator, in effect. Its eigenvalues are observed in four dimensions as the particle masses. Its zero eigenvalues are the massless fermions.

A similar discussion can be made for the Rarita-Schwinger operator. There are many, equivalent ways[6] to write the 4+n dimensional Rarita-Schwinger equation. One way is

$$0 = \Gamma^{\mu} \left(D_{\mu} \psi_{\nu} - D_{\nu} \psi_{\mu} \right) \tag{3}$$

where $\psi_{\mu\alpha}$ is the Rarita-Schwinger field ($\mu = 1 \ldots 4+n$ is a vector index; α is a spinor index). In the gauge $\Gamma^\mu \psi_\mu = 0$, (3) reduces to

$$\begin{aligned} 0 &= \Gamma^\mu D_\mu \psi_\nu \\ &= \not{D}^{(4)} \psi_\nu + \not{D}^{(n)} \psi_\nu \end{aligned} \tag{4}$$

Again zero modes of $\not{D}^{(n)}$ are observed as massless particles in four dimensions. The general zero mode is a sum of modes of two special kinds. For $\nu = 1 \ldots 4$, $\not{D}^{(n)}$ is the ordinary spin 1/2 Dirac operator, and the zero modes are spin 3/2 fermions in four dimensions. For $\nu = 5 \ldots 4{+}n$, the zero modes have spin 1/2 as seen in four dimensions, while their dependence on the compact dimensions is determined by the gauge condition *and* the Dirac-like equation:

$$\begin{aligned} 0 &= \sum_{\nu=5}^{4+n} \Gamma^\nu \psi_\nu \\ 0 &= \sum_{\mu=5}^{4+n} \Gamma^\mu D_\mu \psi_\nu \end{aligned} \tag{5}$$

These conditions, taken together, are equivalent to the gauge invariant internal Rarita-Schwinger equation

$$\begin{aligned} &\sum_{\mu=5}^{4+n} \Gamma^\mu (D_\mu \psi_\nu - D_\nu \psi_\mu) = 0 \\ &\nu = 5 \ldots 4{+}n \end{aligned} \tag{6}$$

in a particular gauge. (We temporarily introduced a gauge fixing condition only to decouple the Minkowski dimensions from the compact ones.) We see that zero modes of the internal Rarita-Schwinger operator become massless spin 1/2 fermions in four dimensions.

Now, can either of these operators have zero eigenvalues? And can the zero eigenvalues form complex representations of a symmetry group?

The crudest problem, which was pointed out in (4), arises in an odd number of dimensions. For odd n, the group O(n) only has one spinor representation. Likewise, the group O(1, 3+n) has only one spinor representation, which transforms under O(1,3) × O(n) as the product of the four component spinor of O(1,3) with the unique spinor of O(n). This being so, fermions that are left or right handed in four dimensions transform the same way under transformations of the internal space. They obey the same Dirac equation in the internal space (modulo non-minimal terms which cannot affect the quantum numbers) so they have the same quantum numbers and furnish a real repesentation of any relevant symmetry group.

In an even number of dimensions the situation is more subtle. For even n the operator $\hat{\Gamma} = \Gamma^1 \ldots \Gamma^n$ anticommutes with all Γ^i, so it is a c-number in any representation of the Clifford algebra. Since $\hat{\Gamma}^2 = \pm 1$ (depending on n) the representation space of the Clifford algebra decomposes into two eigenspaces of $\hat{\Gamma}$, the eigenvalues being ±1 or ±i. Since $\hat{\Gamma}$ commutes with the O(n) group generators $1/4\,[\Gamma^i, \Gamma^j]$, the group has two inequivalent spinor representations, labeled by the eigenvalue of $\hat{\Gamma}$.*

In a world of 4+n dimensions we define

$$\bar{\Gamma} = \Gamma^1\Gamma^2 \ldots \Gamma^{4+n}$$

$$\Gamma^{(4)} = \Gamma^1\Gamma^2 \ldots \Gamma^4$$

$$\Gamma^{\mathrm{Int}} = \Gamma^5\Gamma^6 \ldots \Gamma^{4+n} \qquad (7)$$

---

*For odd n, $\hat{\Gamma}$ commutes with the Γ^i and is a c-number in an irreducible representation of the Clifford algebra. The Clifford algebra thus has two inequivalent representations, labeled by the sign of $\hat{\Gamma}$. They are related to each other by $\Gamma^i \to -\Gamma^i$ (which for odd n yields $\hat{\Gamma} \to -\hat{\Gamma}$), and they are equivalent as representations of O(n) since under $\Gamma^i \to -\Gamma^i$ the group generators $1/4[\Gamma^i, \Gamma^j]$ are unchanged.

These operators have simple interpretations. $\bar{\Gamma}$ labels the spinor representations of O(1, 3+n). $\Gamma^{(4)}$ measures the helicity of four dimensional fermions. And $\Gamma^{\rm Int}$ lables the spinor representations of O(n); it measures what might be called the internal helicity. These operators obey the simple relation

$$\bar{\Gamma} = \Gamma^{(4)} \cdot \Gamma^{\rm Int} \tag{8}$$

This equation has an important consequence. For fixed $\bar{\Gamma}$, the four dimensional and internal chiralities are correlated. If we start with a fermion field restricted to (say) $\bar{\Gamma}=1$ in 4+n dimensions, it breaks down under O(1,3)×O(n) to components with*

$$\eta\Gamma^{(4)} = \eta^{-1}\,\Gamma^{\rm Int} = +1$$

or

$$\eta\Gamma^{(4)} = \eta^{-1}\,\Gamma^{\rm Int} = -1 \tag{9}$$

Fermions of left or right handed physical helicity are left or right handed in the internal space. They obey different Dirac equations, whose zero modes might have different quantum numbers.

This idea quickly runs into trouble if the number of dimensions is divisible by four. In 4k dimensions, $\bar{\Gamma}$ is odd under CPT. This may be seen readily in a Majorana basis, with real gamma matrices. In such a basis CPT acts on spinors just by complex conjugation. $\bar{\Gamma}$ (as defined in (7)) is real in a Majorana basis, but in 4k dimensions it may be readily seen to obey $(\bar{\Gamma})^2 = -1$. The eigenvalues of $\bar{\Gamma}$ are $\pm i$. Being complex conjugates, the eigenvalues of $\bar{\Gamma}$ are related to each other by CPT, and CPT requires that there be equal numbers of fields with $\bar{\Gamma} = +i$ and $\bar{\Gamma} = -i$. Hence there is no net correlation between four dimensional chirality and internal chirality. Fields of $\bar{\Gamma} = +i$ give one correlation and fields of $\bar{\Gamma} = -i$ give the opposite correlation. Thus, in 4k dimensions, CPT requires that the gravitational interactions be vector-like. Naturally, therefore, if the weak interactions are part of the gravitational force in 4+n dimensions, the weak interactions are also vector-like. Alternatively, one may say[13,18] that in 4k dimensions, bare masses are possible for any fermions coupled to gravity only. Such a theory will, of course, always reduce to an appropriate four dimensional theory in which bare masses are still possible.

---

*Here η is a phase factor that will be determined momentarily. It can be ignored for the time being.

In 4k+2 dimensions, the situation is very different. In this case $\bar{\Gamma}^2 = +1$ so $\bar{\Gamma}$ has eigenvalues ±1. CPT leaves $\bar{\Gamma}$ unchanged, and we can consider a theory with fermions of (say) $\bar{\Gamma} = +1$ only. (This option is forced on us in certain situations, for instance in certain ten dimensional supersymmetric field theories and string theories.) This corresponds roughly to a theory with V-A gravitational interactions that forbid fermion bare masses; the question is whether V-A gravity can reduce to V-A weak interactions in four dimensions.

The special role of 4k+2 dimensions in multi-dimensional field theory was first raised in constructing supersymmetric Yang-Mills theories.[19] In the context of analyzing fermion quantum numbers this point was made and developed in references (13) and (14). Similar observations are important in grand unified model-building.[9] In the mathematical literature the periodicity of the spinor representation is an old observation.[20]

Let us now work out the phase factor η of equation (9). We note that as defined $\Gamma^{(4)}$ is (in a Majorana basis) a real matrix whose square is -1, so the eigenvalues of $\Gamma^{(4)}$ are $\pm i$. In 4k+2 dimensions Γ^{Int} likewise has square

-1 and eigenvalues $\pm i$. (In 4k dimensions $(\Gamma^{\mathrm{Int}})^2 = +1$.) A fermi field that obeys $\bar{\Gamma} = \Gamma^{(4)} \cdot \Gamma^{\mathrm{Int}} = +1$ therefore has

$$\begin{aligned} &(A) \quad \Gamma^{(4)} = -\Gamma^{\mathrm{Int}} = +i \\ &\qquad\qquad \text{or} \\ &(B) \quad \Gamma^{(4)} = -\Gamma^{\mathrm{Int}} = -i \end{aligned} \tag{10}$$

A CPT transformation will complex conjugate the eigenvalues, so eigenvalues of type (A) and (B) are exchanged by CPT. This is as it should be . A zero mode of the internal Dirac or Rarita-Schwinger operator with $\Gamma^{\mathrm{Int}} = -i$ corresponds to a left handed massless fermion in four dimensions. Its complex conjugate will have $\Gamma^{\mathrm{Int}} = +i$ and corresponds to a right handed massless fermion in four dimensions. Massless fermions in four dimensions will transform in a complex representation of some symmetry group G if the zero modes of the internal Dirac operator with $\Gamma^{\mathrm{Int}} = -i$ form a complex representation of G, or equivalently, if the zero modes of $\Gamma^{\mathrm{Int}} = +i$ transform differently from those of $\Gamma^{\mathrm{Int}} = -i$.

Since the remainder of this section and the next one will deal with "chiral theories of gravity" with elementary fermi fields of a definite value of $\bar{\Gamma}$, it should be mentioned at the outset that these theories suffer from a major problem. The fermion one loop diagrams in an external gravitational field are anomalous.[18] (The anomaly first appears in a diagram with 4k+2 external gravitons.) In a few special theories in six or ten dimensions the anomalies cancel between fields of different spin; beyond ten dimensions this is impossible. We will not limit ourselves to the anomaly free theories but will investigate the fermion quantum numbers that emerge (at the tree level) in the whole class of chiral gravity theories. There are several justifications for ignoring the anomalies. First, general methods for treating the whole class of chiral gravity theories are as simple as any special methods for analyzing the particular anomaly free theories. And the general methods are likely to be important for other attempts to calculate fermion quantum numbers in Kaluza-Klein theory; for instance, we shall apply them from a different standpoint in section VI. Second, it may happen that in the future a massless field of some exotic spin might be successfully coupled to gravity. This could expand the room for cancellation of anomalies without affecting our tree level considerations for spin 1/2 and spin 3/2 fields. Third, though it seems unlikely, perhaps there is some way to make sense of anomalous field theories or of other theories whose low energy limit is an anomalous field theory.

A rather simple argument due essentially to Lichnerowicz[21] severely limits the possibility of obtaining zero modes of the Dirac operator in complex represenations of a symmetry group. If one squares the internal Dirac operator (which will simply be denoted $i\not{D}$; we henceforth suppress the four Minkowskian dimensions) we find

$$
\begin{aligned}
(i\not{D})^2 &= -\Sigma D_i D^i - \frac{1}{4}\left[\gamma^i,\gamma^j\right]\left[D_i,D_j\right] \\
&= -\Sigma D_i D^i - \frac{1}{32}\left[\gamma^i,\gamma^j\right]\left[\gamma^k,\gamma^\ell\right] R_{ijk\ell} \\
&= -\Sigma D_i D^i + \frac{1}{4} R \qquad (11)
\end{aligned}
$$

Since $-\Sigma D_i D^i$ is a non-negative operator, this shows that if $R > 0$ everywhere the Dirac operator has no zero eigenvalues. Of course, in (11) we have considered a minimally coupled Dirac operator. If non-minimal couplings are present, the Dirac operator may have zero eigenvalues. But the fact that the minimally coupled operator has no zero eigenvalues at all, and leads to no massless fermions in four dimensions, means that even in the presence of minimal couplings, the zero eigenvalues form real representations of whatever symmetry group may be present.* A particularly important case of this is the following. Suppose the compact space B has a symmetry group G. In general there will be many G-invariant metrics on B. If even one of them has $R > 0$, then for any G-invariant metric on B, the zero modes of the Dirac operator, if any, form a real representation of G. (Of course, under these circumstances the Dirac operator will generically have no zero eigenvalues.)

Much of the literature on Kaluza-Klein theory has concerned homogeneous spaces $B = G/H$, G and H being compact non-abelian groups. These spaces all admit a canonical G-invariant metric of positive scalar curvature, so (even if non-minimal terms are added or a different G-invariant metric is used) they give real representations for zero eigenvalues. More generally, a theorem by Lawson and Yau[22] shows that on any compact space B (not necessarily a homogeneous space) with a non-abelian symmetry G, there is a G-invariant metric of positive scalar curvature R. For non-abelian groups such as $SU(3) \times SU(2) \times U(1)$, this rules out the possibility of getting zero modes of the Dirac operator in complex representations.

This simple line of argument does not address the question (of conceptual but probably not of practical interest) of whether zero modes of the Dirac operator can form a complex representation of an abelian symmetry group. (Manifolds with a continuous abelian symmetry group in general do not admit an invariant metric of positive scalar curvature.) Much more important, this line of reasoning does not extend to the Rarita-Schwinger operator whose square is more complicated than (11) and is not manifestly positive even if $R > 0$. For this reason, the Rarita-Schwinger operator can have zero eigenvalues more readily than the Dirac operator. For instance, in four dimensions there is one compact manifold that is not flat but obeys $R_{\mu\nu} = 0$. It is the K3 surface, and for topological reasons it has two zero eigenvalues of the Dirac operator and 42 zero eigenvalues of the Rarita-Schwinger operator.

---

* This should be obvious "physically" from the connection with fermion quantum numbers in four dimensions. The precise mathematical argument will be given shortly.

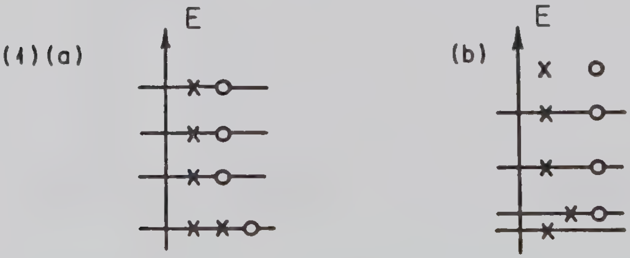
In multi-dimensional supergravity and superstring theories -- which are the only known theories in which fermions are really unified with gravity -- we inevitably are dealing with Rarita-Schwinger fields. It therefore is important to learn to analyze the zero modes of these fields.

There is another no-go theorem, due to Atiyah and Hirzebruch,[15] which for our particular problem is much more restrictive than the reasoning sketched above. They proved precisely that for any continuous symmetry group, abelian or non-abelian, the Dirac zero modes form a real representation. As we will see, their argument has important implications for the Rarita-Schwinger case; for instance, we will use it to prove that if the compact space is a homogeneous manifold, the Rarita-Schwinger operator always leads to a real representation. We will present in section IV an elementary proof of the Atiyah-Hirzebruch theorem that is closely related to the original argument.

Why would a wave operator have zero eigenvalues? And why would these zero eigenvalues form complex representations? We will illustrate the relevant concepts in terms of the Dirac operator. In the rest of this paper, we suppress the four Minkowskian dimensions and concentrate on properties of the n dimensional Kaluza-Klein space B. To streamline notation, gamma matrices are henceforth gamma matrices $\Gamma^1, \Gamma^2, \ldots, \Gamma^n$ of B, and we define an operator $\hat{\Gamma} = i\Gamma^{\mathrm{Int}} = i\Gamma^1\ldots\Gamma^n$ with eigenvalues ±1. Indices i, j, k refer to the internal space; indices μ, ν, α refer to all 4+n dimensions.

Let us define a "Hamiltonian" $H = (i\not{D})^2$. Since $[\hat{\Gamma}, H] = 0$, H eigenstates can be chosen to be at the same time $\hat{\Gamma}$ eigenstates. If $H\psi = E\psi$, then $H\cdot i\not{D}\psi = E\cdot i\not{D}\psi$. So ψ and $i\not{D}\psi$ are degenerate in energy, unless $i\not{D}\psi = 0$. But since $\not{D}\hat{\Gamma} = -\hat{\Gamma}\not{D}$, ψ and $i\not{D}\psi$ have opposite eigenvalues of $\hat{\Gamma}$. Consequently (figure (1)), the H eigenvalues of non-zero energy are paired. For every state of $\hat{\Gamma} = 1$ there is a state of $\hat{\Gamma} = -1$. The zero eigenvalues need not be paired in this way. The number of zero eigenvalues of $\not{D}$ with $\hat{\Gamma} = 1$ minus the number with $\hat{\Gamma} = -1$ is called the index of $\not{D}$.

The index is invariant under arbitrary deformations of $\not{D}$ that preserve the property $\hat{\Gamma}\not{D} = -\not{D}\hat{\Gamma}$, since no smooth distortion of figure (1) which pre-



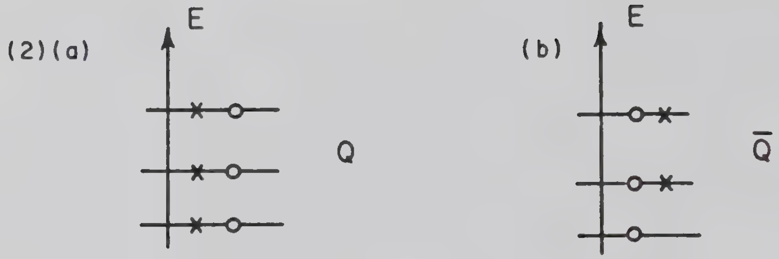
(1) Zero modes of the Dirac operator of positive or negative chirality are denoted by x or o respectively. The number of x's minus o's at zero energy is invariant under perturbations.

serves the pairing at non-zero energy can disturb whatever chirality imbalance may exist at $E = 0$. In particular the index of $\not{D}$ is a topological invariant, depending on the topology of B but not on its metric tensor. Generically, in the absence of some symmetry principles (which can, however, change the situation, as we will see), zero eigenvalues of $\not{D}$ all have positive chirality or all have negative chirality. This is so because zero eigenvalues of equal and opposite chirality would gain non-zero energy under a generical perturbation.

Although the index is the simplest deformation invariant of the Dirac operator that is relevant to the occurrence of zero modes, for our purposes we need a slightly different concept. In 4k+2 dimensions the index of the Dirac operator always vanishes, for the simple reason that the positive and negative chirality zero modes of the Dirac operator are complex conjugates of each other (as we have seen earlier) and therefore equal in number. The concept of interest to us is what mathematicians call the G-index or the character-valued index of the Dirac operator.

Let the manifold B have a symmetry group G. The eigenvalues of H or $i\not{D}$ will then form representations of G. Pick a representation Q, and draw the same picture as before (figure (2)), but only counting multiplets in the Q representation. We define $\text{index}_Q(\not{D})$ to be the number of zero mode multiplets in the Q representation of positive chirality minus the number of zero mode multiplets in the Q representation of negative chirality. For reasons similar to those given earlier, $\text{index}_Q(\not{D})$ is invariant under arbitrary perturbations that respect G symmetry and preserve the fact that $\not{D}\hat{\Gamma} = -\hat{\Gamma}\not{D}$.

Of course, we can still complex conjugate our eigenstates of H. This still reverses the eigenvalue of $\hat{\Gamma}$, but now it exchanges Q with its complex



(2) An x or o now indicates a positive or negative chirality multiplet in the Q or $\bar{Q}$ representation ((a) and (b) respectively). In passing from (a) to (b) the x's and o's are exchanged.

conjugate representation $\bar{Q}$. By complex conjugation (part (b) of figure (2)), this implies that $\mathrm{index}_Q(i\not{D}) = -\mathrm{index}_{\bar{Q}}(i\not{D})$. Upon reduction to four dimensions, this implies the perfectly valid statement that the number of left handed massless fermions in the Q representation equals the number of right handed massless fermions in the $\bar{Q}$ representation.

An equivalent way to define the character-valued index is as follows. The positive and negative chirality zero modes of $i\not{D}$ form representations Λ^+ and Λ^- of G. For $g \in G$ we define

$$\mathrm{index}(g) = \mathrm{tr}_{\Lambda^+}(g) - \mathrm{tr}_{\Lambda^-}(g) \tag{12}$$

Or equally well we define

$$\mathrm{index}(g) = \sum_Q \mathrm{index}_Q(i\not{D})\ \chi_Q(g) \tag{13}$$

where $\chi_Q(g)$ is the trace of g in the Q representation.*

If the character-valued index is nonzero, $i\not{D}$ must have zero modes. Generically, the number of zero modes will be the minimum required to yield the right value of $\text{index}_Q(i\not{D})$ for each Q. The spectrum of zero modes required by the character-valued index we will call the "stable spectrum" of zero modes. We will usually assume that the actual spectrum of zero modes coincides with the stable spectrum and can be computed by evaluating the character-valued index. From the fact that SU(3)×SU(2)×U(1) forbids bare masses for all the known quarks and leptons, it appears that this assumption is valid in nature. A successful model would be one in which the character-valued index consists of three families minus three anti-families.

Unfortunately, as we will see in section IV, the Atiyah-Hirzebruch theorem ensures that the character-valued index always vanishes for the Dirac operator in theories without elementary gauge fields. At least on homogeneous spaces, this is also true for the Rarita-Schwinger field. It is not true (even on homogeneous spaces) for fields of spin 5/2 or larger; we will discuss a counter example in section V (but there does not seem to be any way to use massless fields of spin ≥5/2 in physics [23]). The character-valued index also need not vanish in the presence of elementary gauge fields, and we will construct some pseudo-realistic models in section VI on the basis of this fact.

As we will see, there are very powerful methods for calculating the character-valued index of arbitrary operators. It is never necessary to write down and solve an explicit differential equation.

---

*This definition of index(g) makes sense in 4k as well as 4k+2 dimensions, though it is then not related to our physical problem, and we lose the identity $\text{index}(g) = (\text{index}(g^{-1}))^*$ that follows from complex conjugation in 4k+2 dimensions.

III. OPERATORS ON HOMOGENEOUS SPACES

For our first experience in calculating the character-valued index of various operators, we will consider the simple case in which the manifold B is a homogeneous space G/H. In that case, there is a particularly elementary way [24] to compute the stable spectrum of zero modes of the Dirac operator (or any other G-invariant operator). One simply expands the spinor fields on B in harmonics (irreducible representations) of the group G. For any representation Q of G, let $n_+(Q)$ be the number of times the Q representation appears in the harmonic expansion of positive chirality spinors, and let $n_-(Q)$ be the number of times the Q representation appears in the expansion of negative chirality spinors. By standard theorems about homogeneous spaces and elliptic operators, $n_+(Q)$ and $n_-(Q)$ are always finite, and $n_+(Q) = n_-(Q)$ for all but finitely many Q. Moreover, $\mathrm{index}_Q(i\not{D}) = n_+(Q) - n_-(Q)$. After all, the index is the difference between the number of positive and negative chirality multiplets. This difference normally must be regularized since there are infinitely many states in Hilbert space, and the regularization usually involves pairing off the states of equal, non-zero energy. In a homogeneous space, working in a subspace of Hilbert space defined by a definite representation Q reduces the problem to a finite dimensional problem. No regularization is needed; $\mathrm{index}_Q(i\not{D})$ is just the difference $n_+(Q) - n_-(Q)$ between the number of positive chirality and negative chirality Q multiplets.

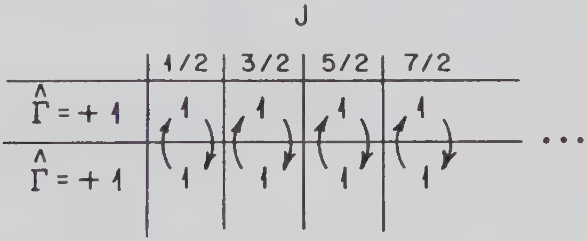
Let us illustrate these ideas with some simple examples. Consider first a particle of spin 1/2 moving on the ordinary two dimensional sphere S^2. Let $\vec{J}$ be the angular momentum operator, and define the "helicity" of a particle at $\vec{x}$ to be the component of angular momentum about the $\vec{x}$ axis. It equals $\pm 1/2$; it is $+1/2$ for states of positive chirality, $-1/2$ for states of negative chirality.

It is well known that a particle of helicity h can be in a state of total angular momentum $J=|h|$, $|h|+1$, $|h|+2$, ... with each allowed value of J appearing exactly once in the harmonic expansion. This is why J=1 is the lowest possible value for photons, and J=2 is the lowest possible value for gravitons.

Whether the chirality is positive or negative, the absolute value of the helicity of a spin 1/2 particle is 1/2. So the allowed values of angular momentum are the same for each chiralty:

$$\begin{aligned} \hat{\Gamma} = +1 &: \quad J = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots \\ \hat{\Gamma} = -1 &: \quad J = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots \end{aligned} \tag{14}$$

Now we can see that the Dirac operator on the sphere has no stable spectrum of zero modes. Since the Dirac operator commutes with J but reverses chirality, acting on (say) the multiplet of given J and chirality ±1, the Dirac operator gives either zero or else the multiplet of the same J and chirality ∓1 (see figure (3)). Since the Dirac operator is hermitian, it either ex-



(3)

(3) The angular momentum and chirality spectrum of the Dirac operator on a sphere. The Dirac operator permutes the states in the way indicated by the arrows.

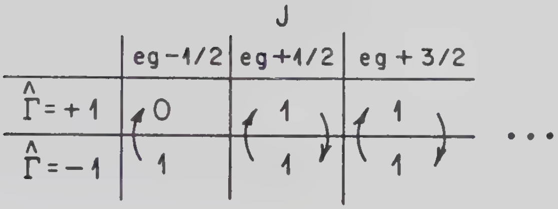
changes these two multiplets or annihilates both of them. Therefore the positive chirality zero modes have the same eigenvalues of J as the negative chirality zero modes, and the character valued index vanishes.

This result could be obtained in various other ways. It follows from the fact that the rotation group SU(2) has no complex representations, or from the fact that a reflection of the two sphere reverses parity and exchanges the two chiralities, or from the fact that the two sphere with its usual metric has positive scalar curvature so that (by Lichnerowicz's theorem) the Dirac operator has no zero modes at all. Now, however, we will consider a slightly modified problem in which the character-valued index is non-zero.

Place at the center of the sphere a magnetic monopole of strength $eg = n/2$, for some integer n. The angular momentum operator now acquires an extra piece[25] $eg\hat{x}$ related to the quantization of magnetic charge. This adds eg to the fermion helicity, so that a fermion of chirality +1 has effective helicity eg+1/2 and a fermion of chirality -1 has effective helicity eg-1/2. If, say, eg>0, the allowed values of angular momentum are now

$$\begin{aligned} \hat{\Gamma} = +1 :& \quad J = eg + \frac{1}{2},\ eg + \frac{3}{2},\ eg + \frac{5}{2},\ \ldots \\ \hat{\Gamma} = -1 :& \quad J = eg - \frac{1}{2},\ eg + \frac{1}{2},\ eg + \frac{3}{2},\ eg + \frac{5}{2},\ \ldots \end{aligned} \tag{15}$$

The crucial point is now that states of $J = eg - \frac{1}{2}$ exist for chirality -1 but not for chirality +1. The Dirac operator must annihilate these states, because acting on states of $\hat{\Gamma} = -1$, $J = eg - \frac{1}{2}$, the Dirac operator would give states of $\hat{\Gamma} = +1$ and $J = eg - \frac{1}{2}$, and such states do not exist. Other multip-



(4)

(4) The quantum numbers of a charged Dirac particle on a sphere in the presence of a magnetic monopole field.

---

lets cancel out as before (figure (4)), so the stable spectrum of zero modes is a single multiplet of $\hat{\Gamma} = -1$ and $J = eg - \frac{1}{2}$.

We will obtain this answer in a different way in section IV as an illustration of a much more general and powerful method of calculating the character-valued index of an operator. The technique for harmonic expansions for G-invariant operators on a homogeneous space G/H does not seem to be well known among physicists. It is explained, for instance, in Appendix IV of Salam and Strathdee, reference (3). In later sections we will occasionally state without detailed derivation results obtainable form harmonic expansions.

IV. THE ATIYAH-HIRZEBRUCH THEOREM

We now turn to the proof of the Atiyah-Hirzebruch theorem, which states that the character-valued of the Dirac operator vanishes on any manifold with a continuous symmetry group (in any even number of dimensions, though our main interest is $d = 4k+2$). The presentation will parallel a recent treatment of Morse theory,[16] and is essentially a more concrete version of the original proof.

Let B be a compact Riemannian manifold of even dimension n. Suppose B admits the action of a symmetry group G. We wish to prove that index(g) = 0 for every $g \in G$. Since every element of G can be approximated arbitrarily well by elements of suitably chosen U(1) subgroups of G, it suffices to prove that index(h) = 0 whenever h is an element of any U(1) subgroup R of G. We therefore specialize to the case of a U(1) symmetry group R. Since the representations of $R \cong U(1)$ are labeled by an integer or half-integer n,* it suffices to show that $\mathrm{index}_n(i\not{D}) = 0$ for all n.

Let ϕ^i be a local coordinate system of B. Let $K^i(\phi^j)$ be the Killing vector field that generates R. (This means that, infinitessimally, the R transformation is $\phi^i \to \phi^i + \varepsilon K^i(\phi^j)$.) Acting on spinors, the generator of R is the "Lie derivative" operator:

$$\mathcal{L}_K = i\left(K^i D_i + \frac{1}{4} \Gamma^{ij} \left(D_i K_j\right)\right) \tag{16}$$

where $\Gamma^{ij} = \frac{1}{2} \left[\Gamma^i, \Gamma^j\right]$. Using the Killing vector equation $D_i K_j + D_j K_i = 0$ and standard identities, it is not difficult to verify that $\mathcal{L}_K$ and $i\not{D}$ commute -- as should be the case since $\mathcal{L}_K$ generates a symmetry. Therefore we may study simultaneous solutions of the equations

---

*The eigenvalues on spinor states are half integers in certain cases.

$$\mathcal{L}_K\psi = n\psi$$
$$i\not{D}\psi = \lambda\psi \tag{17}$$

Our problem is to show that the Dirac index $\mathrm{index}_n(i\not{D})$ vanishes for each sector of Hilbert space labeled by the integer (or half-integer) n.

The basic property of this index is that it is invariant under arbitrary deformations of the operator $i\not{D}$ that are U(1) invariant and preserve the property $i\not{D}\hat{\Gamma} = -\hat{\Gamma}i\not{D}$. Let us therefore perturb the Dirac operator in a way that preserves these properties and simplifies the analysis of its spectrum. Instead of $i\not{D}$ we will study

$$i\not{D}_t = i\not{D} + t\ \Gamma^i K_i \tag{18}$$

where t is a conveniently chosen real number. The character valued index of $i\not{D}_t$ must be independent of t. We will prove that $\mathrm{index}_n(i\not{D})$ vanishes for $n>0$ by studying the behavior as $t \to +\infty$; and we will prove that $\mathrm{index}_n(i\not{D})$ vanishes for $n<0$ by studying the behavior as $t \to -\infty$.

We define a "Hamiltonian"

$$H_t = (i\not{D}_t)^2 = (i\not{D})^2 + t^2K^2 + 2it\ K^jD_j + it\Gamma^{ij}\ D_iK_j \tag{19}$$

If we could show that for sufficiently large t, H_t has no zero eigenvalues, this would establish the vanishing of the character valued index. The general reason this might be true is that the t^2K^2 term is positive definite and becomes very large for large t. However, the analysis is made subtle by the term $2it\ K^jD_j$, which is not positive definite and can have large matrix elements.

A crucial observation is that in the sector $\mathcal{L}_K\psi = n\psi$, H_t reduces to

$$H_t^{(n)} = (i\not{D})^2 + t^2K^2 + 2tn + \frac{1}{2}\, it\, \Gamma^{ij}\, D_iK_j \qquad (20)$$

Since t is freely at our disposal, we choose $t>0$ if $n\geq 0$ and $t<0$ if $n<0$. In this way the term 2tn, as well as the t^2K^2 term, is positive.

Now, if the Killing vector field K^i has no zeros, then all eigenvalues are of order t^2 as $t \to \infty$. In this case, the character valued index certainly vanishes. In general, however, K^i vanishes at certain points, and our analysis is more difficult.

For large $|t|$, the spectrum of $H_t^{(n)}$ can be calculated in an asymptotic expansion in powers of $1/|t|$ by expanding near the minima of the potential. The relevant minima (which might give states that do not diverge in energy as $|t| \to \infty$) are zeros of K. For simplicity, we will treat the case of an isolated zero of K, but the general case is not much different.

We may take our isolated zero of K to be at $\phi^i=0$. Near $\phi^i=0$ we can choose the locally Euclidean coordinates ϕ^i to be such that $K_i = \omega_{ij}\,\phi^j + O(\phi^2)$, with ω_{ij} a constant matrix

$$\omega_{ij} = \begin{bmatrix} 0 & r_1 & & & & & \\ -r_1 & 0 & & & & & \\ & & 0 & r_2 & & & \\ & & -r_2 & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & r_{d/2} \\ & & & & & -r_{d/2} & 0 \end{bmatrix} \qquad (21)$$

(Here d is the dimension of our manifold; the r_i are integers since K generates a U(1) group.) For large $|t|$, keeping only terms that contribute an amount of order t to the energy, $H_t^{(n)}$ simplifies to

$$H_t^{(n)} = -\sum_{i=1}^{d} \frac{\partial^2}{\partial\phi_i^{\,2}} + \sum_{\ell=1}^{d/2} t^2 r_\ell^{\,2} \left(\phi_{2\ell-1}^2 + \phi_{2\ell}^2\right) + 2tn - t\sum_{\ell=1}^{d/2} \frac{ir_\ell}{2} \left[\Gamma_{2\ell-1}, \Gamma_{2\ell}\right] \qquad (22)$$

The ground state energy of (22) is easily calcualted. The first two terms are a sum of commuting harmonic oscillator Hamiltonians. The matrices $\frac{i}{2}\left[\Gamma_{2\ell-1}, \Gamma_{2\ell}\right]$ commute with each other and with the rest of the Hamiltonian, and have eigenvalues ± 1. The ground state energy of (22) is $|t| \sum_{i=1}^{d/2} |r_i| + 2tn$.* This is a good approximation for large $|t|$. Considering $t \to +\infty$ if $n>0$, and $t \to -\infty$ if $n<0$, we see that $H_t^{(n)}$ has no zero eigenvalues and consequently that the character-valued index of $i\not{D}$ vanishes.

If the zeros of K are not isolated points, the discussion must be changed only slightly. Let F be any connected component of the submanifold on which K vanishes. The potential t^2K^2 vanishes on F, so our spin 1/2 particle moves freely along F, but the motion orthogonal to F is restricted to a distance of order $1/\sqrt{t}$. The orthogonal motion is governed by an operator similar to (20), and the zero point energy of the orthogonal motion is strictly positive if $|t| \to \infty$ with $tn>0$, again showing that the character valued index vanishes. (The discussion of degenerate Morse theory in reference (16) is similar.) This completes the proof of the Atiyah-Hirzebruch theorem.

---

*If we restrict ourselves to states of $\mathcal{L}_K\psi = n\psi$ (as we should) the ground state energy of (22) is even larger. This of course does not change the conclusion.

Let us, however, now look at the preceding formulas from a different viewpoint, the goal being to obtain the fixed point formula associated with the Atiyah-Singer index theorem.[26] As we will see, this formula is a powerful tool for computing the character valued index when it is not zero.

Let us now study H_t for $t \to +\infty$. Low-lying eigenvalues of H_t are concentrated near zeros of K, which for simplicity we will take to be isolated points. Most of these states have energy of order t, but some have energy that vanishes as $t \to \infty$. (This has been obscured in the presentation until now.) Let $a^{(i)}_{n,\pm}$ be the number of state ψ concentrated near the ith zero whose energy does not diverge as $t \to +\infty$ and which obey $\hat{\Gamma}\psi = \pm\psi$ and $\mathcal{L}_K \psi = n\psi$. (We have proved $a^{(i)}_{n,\pm} = 0$ for $tn > 0$.) Define the "local index" of the Dirac operator at the i^{th} zero as

$$f_i(\theta) = \sum_n e^{in\theta} \left(a^{(i)}_{n,+} - a^{(i)}_{n,-}\right) \tag{23}$$

The character-valued index $I(\theta)$ is obtained as the sum of the local indexes

$$I(\theta) = \sum_i f_i(\theta) \tag{24}$$

since this sum includes the contribution of all states whose energy is not of order $|t|$, and only such states can contribute to $I(\theta)$. Of course, we have proved that the Dirac case has $I(\theta) = 0$, so (24) is a set of restrictions on the $a^{(i)}_{n,\pm}$. However, we will obtain a formula like (24) for other problems in which $I(\theta) \neq 0$.

To compute the $a^{(i)}_{n,\pm}$, we could simply diagonalize the harmonic oscillator Hamiltonian (22). A more efficient method is as follows. Zero energy states must obey $i\not{D}_t \psi = 0$. Here $i\not{D}_t = i\Gamma^j \left(D_j - itK_j\right)$ has a very simple interpretation; it corresponds to a particle interacting with the abelian vector potential $A_j = tK_j$ as well as the metric of the curved manifold under study. For large t, we know the low-lying states are concentrated near zeros of K. If near a zero at (say) $\phi=0$, $K^i = \omega^{ij}\phi_j$, with ω^{ij} a constant matrix, then the "magnetic field" $F_{ij} = \partial_i A_j - \partial_j A_i$ is just the constant matrix $t\omega_{ij}$. So as $t \to \infty$, our problem reduces to the study of the Dirac equation in a constant magnetic field. As in the usual three dimensional case, the ground state energy is zero (but states of zero energy have $n<0$ if $t \to +\infty$ or $n>0$ if $t \to -\infty$).

With ω_{ij} in the canonical form (21), the analysis is very simple. Define

$$i\not{D}_t^{(1)} = i \sum_{i=1,2} \Gamma^i \left(D_i + itK_i\right)$$

$$iD_t^{(2)} = i \sum_{i=3,4} \Gamma^i \left(D_i + itK_i\right)$$

$$\vdots$$

$$iD_t^{(n/2)} = i \sum_{i=n-1,n} \Gamma^i \left(D_i + itK_i\right) \qquad (25)$$

Then $\left(i\not{D}_t\right)^2 = \sum_{j=1}^{n/2} \left(i\not{D}_t^{(j)}\right)^2$ so a solution of $i\not{D}_t\psi = 0$ obeys simultaneously $i\not{D}_t^{(1)}\psi = \ldots = i\not{D}_t^{(n/2)}\psi = 0$. Thus, we need only the well known solution of the problem of a constant magnetic field in two dimensions.

Choose a basis of gamma matrices $\Gamma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\Gamma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$. If the "rotation angle" is r, so $K_1 = r\phi_2$, $K_2 = -r\phi_1$, then the two dimensional Dirac operator is

$$\not{D}_t = \begin{pmatrix} 0 & \left(\partial_1 - tr\phi_1\right) - i\left(\partial_2 - tr\phi_2\right) \\ \left(\partial_1 + tr\phi_1\right) + i\left(\partial_2 + tr\phi_2\right) & 0 \end{pmatrix} \qquad (26)$$

With $\psi = \binom{u}{v}$, $\not{D}_t \psi = 0$ if

$$u(\phi_1, \phi_2) = (\phi_1 + i\phi_2)^k \exp - tr\left(\phi_1^{\,2} + \phi_2^{\,2}\right)$$

$$v(\phi_1, \phi_2) = (\phi_1 - i\phi_2)^k \exp + tr\left(\phi_1^{\,2} + \phi_2^{\,2}\right)$$

$$k = 0, 1, 2, \ldots \tag{27}$$

The chirality operator is $\hat{\Gamma} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. For $t \to +\infty$, if $tr>0$ only positive chirality zero modes are normalizable. If $tr<0$ only negative chirality zero modes are normalizable.

The symmetry generator is

$$\mathcal{L}_K = i\left(\Gamma^i D_i + \frac{1}{4}\Gamma^{ij} D_i K_j\right)$$

$$= -ir\left(\phi_1 \partial_2 - \phi_2 \partial_1\right) + \frac{r}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{28}$$

We see that on the states (27), $\mathcal{L}_K = r\hat{\Gamma}\,(k + 1/2)$.

The local index in the two dimensional problem would be $h(\theta) = \sum_n e^{in\theta}\left(a_{n,+} - a_{n,-}\right)$ where $a_{n,\pm}$ are the number of zero energy states ψ with $\mathcal{L}_K \psi = n\psi$, $\hat{\Gamma}\psi = \pm\psi$. Using (27) and (28), we see that if $r>0$

$$h(\theta) = e^{i\frac{r\theta}{2}} + e^{i\frac{3r\theta}{2}} + e^{i\frac{5}{2}r\theta} + \ldots$$

$$= \frac{e^{ir\theta/2}}{1 - e^{ir\theta}}$$

$$= \frac{i}{2}\,\frac{1}{\sin\frac{1}{2}r\theta} \tag{29}$$

If $r<0$ we get

$$h(\theta) = -e^{-ir\theta/2} - e^{-i\frac{3}{2}r\theta} - e^{-i\frac{5}{2}r\theta} - \ldots$$

$$= -\frac{e^{-ir\theta/2}}{1-e^{-ir\theta}} = \frac{i}{2}\,\frac{1}{\sin\frac{1}{2}r\theta} \tag{30}$$

It is one of the wonders of analytic functions that these expressions are equal, so we need not worry about the sign of r.

In view of the separation of variables (zero eigenvalues of $\not{\partial}_t$ are zero eigenvalues of each of the $\not{\partial}_t^{(n)}$), we can now easily compute the local index $f_i(\theta)$ at the i^{th} zero. It is just

$$f_i(\theta) = \prod_{a=1}^{n/2}\left(\frac{i}{2}\,\frac{1}{\sin\frac{1}{2}r_a^{(i)}\theta}\right) \tag{31}$$

where $r_a^{(i)}$ is the a^{th} rotation angle at the i^{th} zero of k. The character valued index of the Dirac operator is therefore

$$I(\theta) = \left(\frac{i}{2}\right)^{n/2}\sum_i\prod_a\frac{1}{\sin\frac{1}{2}r_a^{(i)}\theta} \tag{32}$$

which is the fixed point formula. Although we obtained it by taking $t \to +\infty$, an analysis for $t \to -\infty$ leads to the same formula.

Since we know that $I(\theta) = 0$, (32) is a set of very restrictive conditions on the $r_a^{(i)}$. Actually the Atiyah-Hirzebruch theorem is an easy consequence of (32). The right hand side of (32) defines a rational function of $w = e^{i\theta/2}$ which vanishes at $w=0$ and $w=\infty$. This function has no poles.

(Individual terms in (32) have poles at $|w|=1$. These poles must cancel after summing over fixed points, for the following reason. An elliptic operator always has only finitely many zero modes, so $I(\theta)$ has an expansion $I(\theta) = \sum_n a_n e^{in\theta}$ with only finitely many non-zero a_n; therefore $I(\theta)$ is always non-singular for real θ.) A rational function without poles that vanishes at infinity is zero, so $I(\theta) = 0$.

For our purposes the virtue of (32) is that it generalizes easily to other problems. Suppose we wish to study a field $\psi_{\alpha A}$ with a spinor index α and some other index A. For instance, A may be a vector index if we wish to study the spin 3/2 field; or A may be a Yang-Mills index. We wish to calculate the character valued index of an operator acting on $\psi_{\alpha A}$. The physically relevant operator may not actually be the Dirac operator $i\Gamma^j D_j$, but we will assume it differs from the Dirac operator only by irrelevant non-minimal terms.

As before, the character valued index may be computed from the large $|t|$ limit of $i\not{D}_t = i\Gamma^j \left(D_j + itK_j\right)$. The index A only enters in the connection used to define D_j, but the connection is irrelevant as $|t| \to \infty$, as shown by the reduction to a flat space harmonic oscillator problem. So our previous determination of the spectrum is still valid.

What is different is the determination of the quantum numbers of the low-lying states. The symmetry generator is now

$$\tilde{\mathcal{L}}_K = \mathcal{L}_K + Q(\phi^k) \tag{33}$$

with an extra term Q (without derivatives) that acts on the A index. (It is a generalization of the extra angular momentum term for a charge interacting with a magnetic monopole.) Let $Q^{(i)}$ be the value of Q at the i^{th} zero of K.

The states near the i^{th} zero that have approximatley zero energy are still given by (27), but near the i^{th} zero, $\tilde{\mathcal{L}}_k$ is $r\hat{\Gamma}\left(k + \frac{1}{2}\right) + Q^{(i)}$. The effect is very simple. In the sums (29) and (30), one has an extra factor Tr $e^{i\theta Q^{(i)}}$ (the trace being over the A index) so now

$$f_i(\theta) = \left(\frac{i}{2}\right)^{n/2} \mathrm{Tr}\, e^{i\theta Q^{(i)}} \prod_{a=1}^{n/2} \frac{1}{\sin \frac{1}{2} r_a^{(i)} \theta} \qquad (34)$$

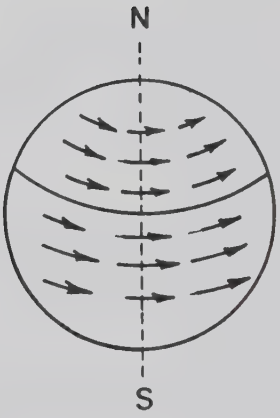
and the fixed point formula is

$$I(\theta) = \left(\frac{i}{2}\right)^{n/2} \sum_i \mathrm{Tr}\, e^{i\theta Q^{(i)}} \prod_{a=1}^{n/2} \frac{1}{\sin \frac{1}{2} r_a^{(i)} \theta} \qquad (35)$$

We will not write down here the more general formula which holds if the zeros of K are not isolated. This formula involves weighting the factors of $1/\left(\sin \frac{1}{2} r\theta\right)$ by the number of zero modes of a certain Dirac operator on the fixed point set.

One may wonder "why" a fixed point formula exists. The character valued index is formally $\mathrm{Tr}\,\hat{\Gamma}\, e^{i\theta\mathcal{L}_K}$ (since states of non-zero energy are paired and cancel out of the trace). The trace of a matrix is the sum of the diagonal matrix elements. In the coordinate basis, the diagonal matrix elements of $e^{i\theta\mathcal{L}_K}$ vanish except near the zeros of K. The fixed point formula is similar to the method of Landau and Lifshchitz[27] for computing the character of a molecular symmetry group furnished by the molecular vibrations in terms of fixed points of the symmetry group action.

To gain some practice with (35), let us use these methods to retrieve the results of section III. We consider a spin 1/2 particle moving on the



(5)

(5) K is taken to generate the rotation of a sphere.

two sphere. We take K to be (figure (5)) the generator of a rotation about the z axis. There are two fixed points, the north pole N and the south pole S. A rotation which is counter-clockwise as seen by an observer looking down at N is clockwise to an observer looking down at S. So the rotation angles are $r=1$ at N and $r=-1$ at S. The fixed point formula gives

$$I(\theta) = \frac{i}{2}\,\frac{1}{\sin\frac{1}{2}\theta} + \frac{i}{2}\,\frac{1}{\sin\left(-\frac{1}{2}\theta\right)} = 0 \tag{36}$$

as expected. Now, as in section III, we assume our spin 1/2 particle to be charged, and we place at the center of the sphere a magnetic monopole with $eg = \frac{n}{2}$, $n \in \mathbb{Z}$. It is well known[25] that the angular momentum operator is shifted: $\vec{J} \to \vec{J} + eg\vec{x}$. In our case $\mathcal{L}_k$ is J_z; the operator Q is the extra piece in J_z or egz. At N, z=1; at S, z=-1. Note that the $Q^{(i)}$ are numbers, ±eg, not matrices, since in the U(1) case the charge index A has only one value. The fixed point formula is now

$$\begin{aligned} I(\theta) &= \frac{i}{2}\, e^{ieg\theta}\, \frac{1}{\sin\frac{1}{2}\theta} + \frac{i}{2}\, e^{-ieg\theta}\, \frac{1}{\sin\left(-\frac{1}{2}\theta\right)} \\ &= -\left(e^{ieg\theta} - e^{-ieg\theta}\right) \frac{e^{-i\theta/2}}{1 - e^{-i\theta}} \\ &= -\sum_{n=-\left(eg-\frac{1}{2}\right)}^{eg-\frac{1}{2}} e^{in\theta} \end{aligned} \tag{37}$$

This agrees with our result from section III, since the sum in (37) is the trace of $e^{i\theta J_z}$ in the represention of $J = eg - \frac{1}{2}$.

The proof of the Atiyah-Hirzebruch theorem that we have given is closely related to the original argument. It has the virtue of yielding the fixed point formula, which has many other applications, as we will see. If one is only interested in the vanishing of the character-valued index of the Dirac operator, the following alternative argument may be sketched. For a space B that admits action of a non-abelian group G, the Lawson-Yau theorem[22] states that B admits a G-invariant metric of positive scalar curvature. (The basic idea of the proof is as follows. Any G-invariant metric g can be decomposed as $g = g_1 + g_2$, where g_1 is the metric transverse to the directions of the group action and g_2 is the metric along the group action. Lawson and Yau show that the metric $g^{\varepsilon} = g_1 + \varepsilon g_2$ has positive scalar curvature if ε is suitably small and positive and g_1 obeys some mild conditions at the fixed points of G.) Combined with the Lichnerowicz theorem (no zero modes of the Dirac operator if the scalar curvature is positive), this implies the vanishing of the character valued Dirac index for manifolds with non-abelian symmetry groups. For manifolds with only an abelian symmetry group, one may reason as follows. Let B be a manifold with U(1) symmetry that violates the Atiyah-Hirzebruch theorem in n dimensions. Let $\tilde{B}$ be a manifold of dimension n+2 defined to be a non-trivial fiber bundle over S^2 with fiber B. (To construct this bundle, consider the space S^3 of pairs of complex numbers $\binom{z_1}{z_2}$ with $|z_1|^2 + |z_2|^2 = 1$. In the product $S^3 \times B$, make the identification $\left(\binom{z_1}{z_2}, \phi^i\right) \approx \left(\binom{z_1 e^{i\alpha}}{z_2 e^{i\alpha}}, e^{in\alpha\mathcal{L}_K} \phi^i\right)$ for suitable n.) The space $\tilde{B}$ has SU(2)×U(1) symmetry, and has a non-trivial character-valued index of the Dirac operator if B does. (To see this, choose on $\tilde{B}$ a Kaluza-Klein metric with large radius in the S^2 directions and tiny radius in the B directions; and solve the Dirac equation on $\tilde{B}$ in a Born-Oppenheimer-Kaluza-Klein approximation.) Since the Lawson-Yau and Lichnerowicz theorems imply that the character-valued index must vanish on $\tilde{B}$, it must also vanish on B.

V. RARITA-SCHWINGER FIELDS

In this section we will study the character valued index of the Rarita-Schwinger field. We will not be able to reach a comprehensive result similar to the Atiyah-Hirzebruch theorem for the Dirac case. We will prove that the zero modes of the Rarita-Schwinger operator on any homogeneous space G/H form a real representation of G. In other words, we wil show that on any homogeneous space of dimension 4k+2, the character valued index vanishes. For homogeneous spaces of dimension 4k, the same argument shows that the character valued index vanishes except for the trivial character (the topological index). Unfortunately, I do not know a general result for the Rarita-Schwinger field on spaces that are not homogeneous spaces. (I also have been unable, despite many attempts, to find a case in which the character valued index is non-trivial, and I believe that if such manifolds exist, they are rather complicated manifolds with rather small symmetry groups in relation to the number of dimensions.) Because our results will not be entirely conclusive, we will return after discussing the theorem on homogeneous spaces to a discussion of various problems in the use of the Rarita-Schwinger operator.

Our basic tool will be the fixed point formula discussed in the previous section. In particular, we will not use the local supersymmetry of the Rarita-Schwinger field; it may be possible to find a stronger result by using this property.

In an appropriate gauge, the Rarita-Schwinger field is simply a vector spinor field $\psi_{\mu\alpha}$ (μ is a vector index, α a spinor index) which obeys, up to irrelevant non-minimal terms, a Dirac equation $\not{D}\psi_\mu = 0$. Of course, we wish to discard zero modes of ψ_μ tht can be gauged away or which violate gauge conditions. Physically, in quantizing a theory, zero modes which are gauge artifacts are canceled by zero modes of the spin 1/2 ghost fields. Therefore, we must subtract from the character valued index of the Rarita-Schwinger field the corresponding index of the spin 1/2 ghosts. (This is the general logic, but actually the ghost index vanishes, by the Atiyah-Hirzebruch theroem.)

Let us work out the fixed point formula for the spin 3/2 field. Consider an isolated zero of the Killing vector field K. Suppose that near the zero (which we assume to be at $\phi^i=0$) $K_i = \omega_{ij}\, \phi^j$ where ω is a constant antisymmetric matrix

$$\omega = \begin{pmatrix} 0 & r_1 & & & & & \\ -r_1 & 0 & & & & & \\ & & 0 & r_2 & & & \\ & & -r_2 & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & r_{n/2} \\ & & & & & -r_{n/2} & 0 \end{pmatrix} \tag{38}$$

The symmetry generator $\tilde{\mathcal{L}}_K$ for a spin 3/2 field is

$$\begin{aligned} \tilde{\mathcal{L}}_K \psi_j &= i\left(K^i D_i \psi_j + \frac{1}{4} \Gamma^{\ell m} \left(D_\ell K_m\right) \psi_j\right) - i\left(D_j\, K^i\right) \psi^i \\ &= \mathcal{L}_K \psi_j - i\left(D_j\, K^i\right) \psi^i \end{aligned} \tag{39}$$

where $\mathcal{L}_K$ is the generator for a spin 1/2 field and the extra piece acting on the vector index is $-i\left(D_j K^i\right)$. At $\phi^i = 0$, $D_j K_i = \omega_{ij}$, so $\tilde{\mathcal{L}}_K$ acts on the vector index of ψ_i by multiplication by the matrix $i\omega$. The factor of $\mathrm{Tr}\, e^{i\theta Q^{(i)}}$ in formula (34) for the local index is here to be replaced by

$$\mathrm{Tr}\, e^{i\theta(i\omega)} = 2 \sum_{a=1}^{n/2} \cos\theta r_a \tag{40}$$

The fixed point formula for the character valued index of the Rarita-Schwinger field is then

$$I(\theta) = \left(\frac{i}{2}\right)^{n/2} \sum_i \left[\left(\prod_{a=1}^{n/2} \frac{1}{\sin \frac{1}{2} r_a^{(i)} \theta} \right) \left(\sum_{b=1}^{n/2} 2\cos r_b^{(i)} \theta - 1 \right) \right] \tag{41}$$

Here i runs over the fixed points or zeros of K; $r_a^{(i)}$, $a = 1 \ldots n/2$ are the rotation angles at the ith zero; and the minus one in the last factor in (41) is chosen to subtract the index of the ghost fields, as discussed earlier. (Minus one equals minus two plus one; for the quantization of ψ_i, there are[6] two ghosts with the same chirality as ψ_i and one of opposite chirality.) (41) has a generalization when the fixed points are not isolated.

(41) places very severe restrictions on the possibility of obtaining a complex representation of Rarita-Schwinger zero modes. (41) shows that $I(\theta)$ is a rational function of $w = e^{i\theta/2}$ with poles only for $|w| = 1$ or at $w=0$ or $w=\infty$. The poles at $|w| = 1$ must cancel upon summing over fixed points, as in the spin 1/2 case. (Note the discussion following (32).) The rational function $I(w)$ must be a constant unless there really are poles at $w=0$ or $w=\infty$. If the largest rotation angle at the ith fixed point is $r_b^{(i)}$ then the contribution of this fixed point to $I(w)$ behaves for $w \to \infty$ or $w \to 0$ as

$$f_i(w) \sim w^{\left(r_b^{(i)} - \sum_{a \neq b} r_a^{(i)}\right)} \qquad (42)$$

There are poles at $w=0$ and $w=\infty$ if and only if the largest rotation angle $r_b^{(i)}$ is bigger than the sum of the others

$$r_b^{(i)} > \sum_{a \neq b} r_a^{(i)} \qquad (43)$$

If (43) is not obeyed for any i, $I(w)$ has no singularities. Even if (43) is obeyed, the poles may vanish in summing over i. It is difficult to satisfy (43) in a multidimensional space with many rotation angles at each fixed point.

If for each i and each b

$$r_b^{(i)} < \sum_{a \neq b} r_a^{(i)}$$

then $I(w)$ is a rational function without poles and bounded for $w \to \infty$, so is a constant. The constant is an integer -- the ordinary or topological index of the Rarita-Schwinger field. If (44) is always a strict inequality,

$r_b^{(i)} < \sum_{a \neq b} r_a^{(i)}$ for all i and b, then $I(w)$ vanishes as $w \to \infty$, so $I(w) = 0$ and the topological index vanishes.

We will use these considerations to prove that the character-valued index of the Rarita-Schwinger field on a homogeneous space always vanishes except possibly for the trivial character (which may appear in the case of 4k dimensions).

The homogeneous space G/H is defined as follows. It is the space of all $g \in G$ with g and gh considered equivalent for any $h \in H$. Because of the equivalence relation, the dimension of G/H equals the dimension of G minus the dimension of H. The space G/H is invariant under $g \to ug$ for any $u \in G$. A fixed point of this transformation is an element g of G such that $ug = gh$ for some $h \in H$; in other words $g^{-1}ug \in H$. If u is not equivalent up to similarity to an element of H, then the symmetry transformation $g \to ug$ has no fixed points.

If the rank of H is less than the rank of G, then the generical generator A of the Cartan subalgebra of G is not equivalent (up to similarity) to any generator of H. Then $e^{i\theta A}$ acts on G/H without fixed points, so the character valued index vanishes. Hence we need only consider the case rank H = rank G.

Let us now make a brief detour. In general, given two spaces M,N the character-valued Dirac and Rarita-Schwinger indexes of M,N and the product $M \times N$ are related by

$$\mathrm{index}_{\mathrm{Dirac}}(M \times N) = \mathrm{index}_{\mathrm{Dirac}}(M) \cdot \mathrm{index}_{\mathrm{Dirac}}(N)$$

$$\begin{aligned} \mathrm{index}_{\mathrm{R.S.}}(M \times N) = {} & \mathrm{index}_{\mathrm{R.S.}}(M) \cdot \mathrm{index}_{\mathrm{Dirac}}(N) \\ & + \mathrm{index}_{\mathrm{Dirac}}(M) \cdot \mathrm{index}_{\mathrm{R.S.}}(N) \\ & + \mathrm{index}_{\mathrm{Dirac}}(M) \cdot \mathrm{index}_{\mathrm{Dirac}}(N) \end{aligned} \tag{45}$$

These equations hold because the Dirac and Rarita-Schwinger equations on M×N can be solved by separation of variables. The second equation in (45) (which is our real interest) arises because the vector index of ψ_i must be tangent to either M or N, so a Rarita-Schwinger solution on M×N obeys the Dirac equation on M and the Rarita-Schwinger equation on N or vice-versa (the last term in the second equation in (45), which is not intuitively obvious, arises in subtracting the ghost contributions from the Rarita-Schwinger indexes). In particular, (45) implies that if M and N have vanishing Dirac index, then M×N has vanishing Rarita-Schwinger index. Since a homogeneous space has vanishing Dirac index (by the Lichnerowicz or Atiyah-Hirzebruch theorems), a product of two homogeneous spaces has vanishing Rarita-Schwinger index.

From these facts it follows that the Rarita-Schwinger index of G/H vanishes unless G is simple. For suppose $G = G_1 \times G_2$ with non-trivial $G_1 \times G_2$. A subgroup of G of maximal rank is then necessarily $H = H_1 \times H_2$ where H_1 and H_2 are maximal rank subgroups of G_1 and G_2 respectively. Then $G/H = (G_1/H_1) \times (G_2/H_2)$ is a product of homogenous spaces, and has vanishing character valued Rarita-Schwinger index.

We still must study G/H with G simple and H a subgroup of maximal rank. Let us first calculate the rotation angles at the fixed points for a typical infinitessimal transformation $g \to (1 + i\varepsilon A)g$, A being a generator of G. Since H is maximal, A is equivalent (by conjugation) to a generator of H, so we may assume A is actually such a generator. A typical fixed point is then $g = 1$; because of the equivalence under right multiplication by an element of h, the transformation $g \to (1 + i\varepsilon A)g$ is equivalent to

$$g \to (1 + i\varepsilon A)\, g\, (1 - i\varepsilon A) \tag{46}$$

This again makes it clear that $g = 1$ is a fixed point. The fixed point is isolated if A is chosen generically. The rotation angles at $g = 1$ may be computed as follows. The Lie algebra $\mathcal{G}$ of G can be decomposed as the Lie algebra $\mathcal{H}$ of H plus an orthogonal complement $\mathcal{K}$: $\mathcal{G} = \mathcal{H} \oplus \mathcal{K}$. Near $g = 1$, the generic element of G/H is $1 + ik$, with $k \in \mathcal{K}$. The transformation (46) acts on k by $k \to k + i\varepsilon\,[A,k]$. The rotation angles at $g = 1$ are therefore just the eigenvalues of A acting on $\mathcal{K}$ by conjugation.

The other fixed points are at points g_i such that $g_i^{-1} A\, g_i \in \mathcal{H}$. The rotation angles are the eigenvalues of $g_i^{-1} A\, g_i$ acting by conjugation on $\mathcal{K}$.

Now we are ready to prove that the Rarita-Schwinger index vanishes on homogeneous spaces, except for the trivial character. Before stating the argument in a general way let us first consider the case that G is SU(N). Let Y_0 be the SU(N) generator

$$Y_0 = \begin{pmatrix} 1/2 & & & & & \\ & -1/2 & & & & \\ & & 0 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \qquad (47)$$

It generates an SU(2) subgroup of SU(N), so its eigenvalues in any representation are integers or half integers. Let $Y = Y_0 + \varepsilon Q$ where Q is a generic SU(N) generator that commutes with Y_0 and ε is a suitably small real number.

Consider the Rarita-Schwinger equation on some space homogeneous under an SU(N) action. Let Λ^+ and Λ^- be the SU(N) representations of positive and negative chirality zero modes. Define $I(\theta) = \mathrm{Tr}_{\Lambda^+}\, e^{i\theta Y} - \mathrm{Tr}_{\Lambda^-}\, e^{i\theta Y}$. In any non-trivial SU(N) representation the biggest eigenvalue of Y_0 is at least 1/2. Hence, if Λ_+ differs from Λ_- by a non-trivial SU(N) representation, the most positive power of $e^{i\theta}$ appearing will be at least $e^{i\theta\alpha}$ where $\alpha = 1/2 +$

$O(\varepsilon)$. (For suitable Q and sufficiently small ε, there is no accidental cancellation of the highest power. This is the only role of Q and ε in the discussion.) This means that with $w = e^{i\theta/2}$, $I(w)$ -- if not a constant -- diverges for $w \to \infty$ at least as $w^{1+O(\varepsilon)}$. Reference to equation (42) shows the necessary condition to achieve this; we need at one of the fixed points of the fixed points P of the transformation generated by Y

$$r_b - \sum_{a \neq b} r_a > 1 \tag{48}$$

where r_b is the largest rotation angle at P and r_a are the other rotation angles at P.

What are the rotation angles of the transformation Y that are not of order ε? They are one, corresponding to the SU(N) generator

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & . & 0 \end{pmatrix} \tag{49}$$

and its adjoint, and 1/2 corresponding to the generators

$$X_\alpha = \begin{pmatrix} 0 & 0 & a_1 & \dots & a_k \\ 0 & 0 & b_1 & \dots & b_k \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \tag{50}$$

and their adjoints. Of course, in general we do not count A and all the X_α; we only count those that are in $\mathcal{K}$ (the complement in G of the H Lie algebra). The only way to obey (48) is to include A but none of the X_α. This means H must be a subgroup of SU(N) that includes all the X_α and their adjoints and all the diagonal generators (since H has maximal rank) but not A. There is no such subgroup of SU(N). So Λ_+ and Λ_- differ at most only by the trivial character.

The same argument goes through with SU(N) replaced by any simple Lie group G. One simply replaces Y_0 by the generator of the Cartan sub-algebra parallel to a root E of maximum length. E plays the role of A; the roots not orthogonal to E play the role of the X_α. The rest of the argument is unchanged.

Thus, we have shown that on any homogeneous space G/H (of dimension 4k or 4k+2) the character valued index of the Rarita-Schwinger field is a constant, a multiple of the trivial character. It is equal simply to the ordinary index, the difference between the total number of right-handed and left-handed zero modes. The ordinary Rarita-Schwinger index certainly vanishes in 4k+2 dimensions but (unlike the ordinary Dirac index) it need *not* vanish on homogeneous spaces of dimension 4k. It equals one on SU(4)/(SU(2)×SU(2)×U(1)), HP^2, and $G_2/O(4)$. This may be seen by the methods of sections III and IV, or from facts in reference (28), where properties of these spaces are described.

Interestingly, for spin greater than 3/2, it *is* possible to obtain complex representations of zero modes on homogeneous spaces. Consider a spin 5/2 field, which we may represent as a tensor spinor $\psi_{ij\alpha}$, i and j being vector indices and α a spinor index (we suppose $\psi_{ij} = \psi_{ji}$, $\psi i_i = 0$). The wave equation $\not{D}\psi_{ij} = 0$ *can* have zero modes in complex representations on homogeneous spaces. For instance, on the six dimensional manifold CP^3 this operator has the following stable spectrum of zero modes: one multiplet of left handed zero modes in the symmetric tensor representation S^{ij}, and one multiplet of right handed zero modes in the complex conjugate representation S^*_{ij}. This may be seen with the methods of sections III and IV.

Of course, physically sensible couplings of a massless spin 5/2 field to gravity do not seem to exist.[23] However, this arises only because the timelike components ψ_{oi} of the tensor-spinor field have the wrong metric; because of the apparent non-existence of spin 5/2 gauge invariance there is no way to cancel or remove them. Purely as a Euclidean equation, with i and j tangent to the positive signature Kaluza-Klein space, the equation $\not{D}\psi_{ij} = 0$ makes perfect sense and has the properties just stated.

Our result about the Rarita Schwinger operator is much less sweeping than the Atiyah-Hirzebruch theorem of the spin 1/2 case. I do not know whether on some spaces that are not homogeneous the Rarita Schwinger operator may have zero modes in complex representations. (I am convinced, from many unsuccessful attempts to find them, that if such spaces exist they are rather complicated. It is difficult to satisfy equation (42)).

There is actually a strategy which might very plausibly lead eventually to a general proof that the character valued index of the Rarita-Schwinger field always vanishes except for the trivial character. The topological index of an operator is a cobordism invariant; this means that it vanishes for any manifold M of dimension n that is the boundary of a manifold of dimension n+1. The character valued index is likewise invariant under equivariant cobordism; this means that if M admits the action of a group G and is the boundary of a manifold of dimension n+1 to which the G action on M can be extended, then the character valued index vanishes for any operator on M. If a set of generators of the U(1) spin bordism ring (the ring of spin manifolds with U(1) symmetry modulo those which are boundaries) were found, our conjecture about the Rarita-Schwinger field could be proved by showing it to hold for all the generators. The mathematical problem of determining a set of generators for the (oriented) U(1) spin bordism ring has not been solved.

However, the analogous problem has been solved for the unoriented[44] and unitary[45] U(1) bordism rings. These rings are generated by very simple spaces (essentially, homogeneous spaces and fiber bundles in which the fiber is a homogeneous space). If the U(1) spin bordism ring is found to be generated by equally simple spaces, it will be possible to use the methods described above to prove (or disprove) the conjecture that the character-valued Rarita-Schwinger index is always a constant.

Because the situation for spin 3/2 fields is not completely clear, some general remarks on the subject may be useful. It is believed that massless spin 3/2 fields can be consistently coupled to gravity only in locally supersymmetric theories. This apparently means[17] that we are limited to eleven dimensions or less. In addition, beyond ten dimensions the chiral Rarita-Schwinger field has one loop anomalies [18] that spoil general covariance and cannot be canceled by the anomalies of any known fields that can be consistently coupled to gravity. For both of these reasons, it appears that six and ten dimensions are the relevant cases for chiral Rarita-Schwinger fields. This corresponds to two or six compact dimensions respectively.

With two compact dimensions, the only manifolds with continuous symmetry are the sphere, torus, and Klein bottle; the first two are homogeneous spaces, and on the last two the continuous symmetries have no fixed points, so on all of them the character-valued Rarita-Schwinger index vanishes. We turn then to the case of six compact dimensions.

Six compact dimensions are unfortunately too few to admit SU(3)×SU(2)×U(1) symmetry.[4] One may be willing to postulate an elementary U(1) gauge field and to try to obtain only SU(3)×SU(2) as the symmetry group of a six manifold.* The unique six manifold with SU(3)×SU(2) symmetry is $CP^2 \times S^2$. This is a homogeneous space to which our theorem applies; the character-valued index could not be non-zero. Even worse, this space does not admit spinors, so the Rarita-Schwinger equation on $CP^2 \times S^2$ cannot be defined.[29]

However, if the U(1) gauge field has a magnetic monopole expectation value on CP^2, spinors can be introduced (this is the so-called spin_c structure); all fermi fields, including the Rarita-Schwinger field, must have non-zero (half-integral) U(1) charges. The non-zero U(1) charge of the spin 3/2 field introduces new anomalies (the mixed gauge-gravity anomalies of reference (18)) which cannot cancel among themselves unless there are many more than two Weyl gravitinos (two is the maximum of any known or conjectured ten dimensional supergravity theory) and which cannot be canceled by anomalies of spin 1/2 fields (because of a different tensor structure). If we ignore this and proceed, we can calculate the spectrum of the Rarita-Schwinger operator on $CP^2 \times S^2$. The SU(3)×SU(2) invariant expectation value of the U(1) field strength on $CP^2 \times S^2$ depends on two "monopole numbers" -- a half integer p on CP^2 (half integer so as to get a spin_c structure), and an arbitrary integer q on S^2. The resulting zero mode spectrum can be computed by the methods of sections III and IV. One obtains non-trivial complex representations, depending on p and q, but these representations have little resemblance to physics and are anomaly-ridden (because the ten dimensional theory with U(1) coupling to the Rarita-Schwinger field is anomalous).

---

*This suggestion was made independently by M. Gell-Mann.

These accumulated difficulties may encourage us to give up on accommodating SU(3)×SU(2) symmetry in six compact dimensions. We may simply try for SU(3) symmetry. The six-manifolds with SU(3) symmetry are quite restricted and can be seen from the fixed point formula to have vanishing Rarita-Schwinger character valued index. It may be possible to accommodate SU(2)×U(1) in ten dimensions and to obtain leptons but not quarks as Rarita-Schwinger zero modes.

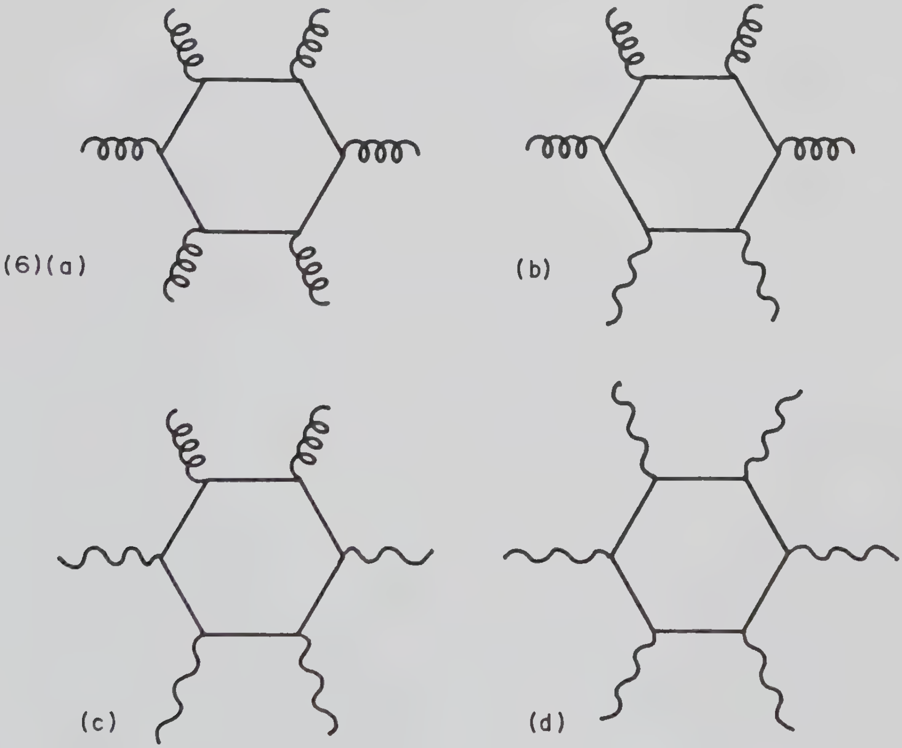
Evidently, whatever is the behavior of the Rarita-Schwinger operator on spaces that are not homogeneous, to obtain physics in this way would not be easy.

VI. ELEMENTARY GAUGE FIELDS

We have so far assumed (except for some brief remarks at the end of the last section) that gauge fields are not elementary but arise as components of the metric tensor in 4+n dimensions. As this assumption has led to difficulties, let us reconsider it. In this section we will suppose that elementary gauge fields do play a crucial role. Thus we are considering a much less ambitious program. We can no longer hope to unify all forces or calculate all couplings, but we still can hope to explain the fermion quantum numbers; we can try to start with a simple fermion representation in 4+n dimensions and obtain a complicated, repetitive one in 4 dimensions after compactification.

We must start in 4+n dimensions with gauge fields that couple differently to left and right handed fermions. For if the gauge quantum numbers of the fermions are vector-like, they will remain vector-like after any compactification.

Non-vectorlike couplings of elementary gauge fields raise an immediate and serious problem of canceling anomalies. In 2p dimensions the diagrams with p+1 external gluons, p-1 external gluons and two gravitons, p-3 external gluons and four gravitons, etc., are all anomalous[18] (figure (6)). If the



(6) The anomalous diagrams in (say) ten dimensions.

gauge field A_μ^a couples to left handed and right handed spin 1/2 fermions via matrices M_L^a and N_R^a, respectively, the condition for anomaly cancellation is

$$\mathrm{Tr}\,\left(M_L^a\right)^r = \mathrm{Tr}\,\left(N_R^a\right)^r$$

$$r = p+1,\ p-1,\ p-3,\ p-5,\ \ldots \tag{51}$$

Equation (51) is trivially obeyed if $M_L^a = N_R^a$, but this is a vector-like theory that will remain vector-like after compactification. Non-trivial solutions of (51) are difficult to find because (51) must hold for many values of r.

If charged spin 3/2 fields are included, the anomaly conditions are more complicated. Anomalous diagrams with four or more external gravitons have different tensor structure for spin 1/2 and spin 3/2 fields and must cancel separately. The anomalous diagrams with zero or two external gravitons (and p+1 or p-1 external gluons) could possibly cancel between spin 1/2 and spin 3/2, in general, but it is impossible to arrange this in known supergravity theories. We will not consider models with charged spin 3/2 fields.

Although it is difficult to find a non-trivial solution of (51), there is a big bonus for doing so. One automatically gets a theory that reduces to an anomaly-free theory in four dimensions, because anomaly free theories remain anomaly free after any compactification. Cancellation of anomalies is the condition under which the effective action in 4+n dimensions is gauge invariant and generally covariant. If it has these properties in 4+n dimensions, it must retain them after any valid approximation, such as approximate reduction to a four dimensional effective action.

A relatively simple solution of (51) is the following. Consider, in 2p dimensions, a theory with gauge group O(2p+6) (or O(2p+4k+6), $k > 0$). Let the positive chirality spinors of the Lorentz group O(1,2p-1) transform as positive chirality spinors of the gauge group O(2p+6); and let the negative chirality O(1,2p-1) spinors transform as negative chirality spinors of O(2p+6). For p=2 this is the usual O(10) model in four dimensions. For any p, this theory is anomaly free. It obeys (51) because the first p+2 Casimir operators of the left and right handed spinor representations of O(2p+6) are equal.* The fermion representation of the O(2p+6) model in 2p dimensions is irreducible in the sense that a combined parity and internal parity operation exchanges the left and right handed fermions. This is the simplest theory with non-trivial anomaly cancellation that I can find in $d > 4$.

Despite starting with a non-vectorlike theory, we are not assured of keeping this property after dimensional reduction. The familiar relation

$$\Gamma_1 \ldots \Gamma_{4+n} = \Gamma_1 \ldots \Gamma_4 \cdot \Gamma_5 \ldots \Gamma_{4+n} \tag{52}$$

shows that the 4+n dimensional chirality operator $\Gamma_1 \ldots \Gamma_{4+n}$ differs from the four dimensional operator $\Gamma_1 \ldots \Gamma_4$ by a factor $\Gamma_5 \ldots \Gamma_{4+n}$ that may be plus or minus one. The non-vectorlike nature is washed out after naive dimensional reduction.

Indeed, if the gauge fields have zero expectation value, they are of no help whatsoever in avoiding the problems discussed in previous sections. They simply do not play any role. But vacuum expectation values of gauge fields that can be smoothly turned on or off are likewise irrelevant, for the usual reasons. We therefore must consider gauge field expectation values that cannot be smoothly turned on or off. The prototype of such a thing is the Dirac magnetic monopole, which because it carries topological information and obeys a Dirac quantization condition cannot be smoothly turned on or off. We must place a generalized monopole "inside" the Kaluza-Klein space. Models of this general type have been considered before by Randjbar-Daemi, Salam, and Strathdee.[14]

---

*This is so because Tr $\sigma_{i_1 j_1} \sigma_{i_2 j_2} \ldots \sigma_{i_k j_k} \cdot \tilde{\Gamma} = 0$ for $k < p+2$, if $\tilde{\Gamma}$ is the product of all 2p+6 gamma matrices of O(2p+6) and $\sigma_{ij} = [\gamma_i, \gamma_j]$ are the group generators.

For illustrative purposes we set p=5 and consider an O(16) theory in 10 dimensions. As explained earlier, the left (or right) handed O(1,9) spinors will be chosen as left (or right) handed spinors of O(16). We will consider three models.

In the first two models we take the six compact dimensions to be CP^3 -- the unique six dimensional space with SU(4) symmetry. We will assume that only a single (abelian) component of the O(16) gauge field has an expectation value.

On CP^3 there is a topologically non-trivial U(1) gauge field. (It is a U(1) connection on the basic non-trivial line bundle over CP^3.) This U(1) gauge field can be chosen to be SU(4) invariant in a unique way. Its strength, like the Dirac monopole charge, must be an integer n in certain units.

Our first two models correspond to two ways of embedding U(1) in O(16).

(i) the embedding

$$\begin{pmatrix} & 1 & & & & & & & \\ -1 & & & & & & & & \\ & & & 1 & & & & & \\ & & -1 & & & & & & \\ & & & & \cdot & & & & \\ & & & & & \cdot & & & \\ & & & & & & \cdot & & \\ & & & & & & & & 1 \\ & & & & & & & -1 & \end{pmatrix} \qquad (53)$$

breaks O(16) to SU(8)×U(1). The four dimensional gauge group is therefore SU(4)×SU(8)×U(1), where SU(4) originates from gravity and SU(8)×U(1) from O(16).

The quantum numbers of fermion zero modes can be computed using the methods of section III or IV. One finds rather complicated anomaly free representations. We will here consider only the minimal case of monopole number $n=1$. The representation of SU(4)×SU(8)×U(1) that emerges is as follows. Let V^i and W^j be the fundamental four and eight dimensional representations of SU(4) and SU(8), respectively. Let S^{ij} be the symmetric product of two V^i, and $\bar{S}_{ij}$ its complex conjugate; and let A^{jk} be the antisymmetric product of two W^j and $\bar{A}_{jk}$ its complex conjugate. Then the left handed zero modes transform as $(V^i, W^j)^3 \oplus (1, A_{ij})^{-2} \oplus (\bar{S}_{ij}, 1)^{-4}$; the superscript is the U(1) charge. The right handed massless fermions transform of course in the conjugate representation $(\bar{V}_i, \bar{W}_j)^{-3} \oplus (1, A^{ij})^2 \oplus (S^{ij}, 1)^4$. Despite its complexity, this representation is anomaly free. For reasons that are not at all clear, this representation is closely related to the supergroup SU(4|8).[30]

(ii) We can instead embed U(1) in O(16) as

$$\left(\begin{array}{ccc|cccccc} 0 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & & & & & \\ \hline & & & & 1 & & & & \\ & & & -1 & & & & & \\ & & & & & & 1 & & \\ & & & & & -1 & & & \\ & & & & & & & & 1 \\ & & & & & & & -1 & \end{array}\right) \tag{54}$$

breaking O(16) to O(10) × SU(3) × U(1). In this case the monopole number must be even, for topological reasons (to ensure proper Dirac quantization for particles in the spinor as well as vector representations of O(10)). So we take the minimal case n=2.

The zero modes can be shown in this case to consist of eight O(10) families (left handed 16 and right handed $\overline{16}$) and no antifamilies. Under SU(4) (from gravity) × SU(3) × U(1) (from O(16)) they transform as follows: they are neutral under SU(3) and transform as $4^3 \oplus \bar{4}^{-3}$ under SU(4) × U(1).

(iii) For our last model, we consider a case in which a non-abelian subgroup of O(16) has a vacuum expectation value. On any six dimensional Riemannian manifold there is a natural O(6) gauge field. One simply takes the spin connection $\omega_\mu{}^i{}_j$ and regards it as a gauge field $A_\mu{}^i{}_j$. We embed O(6) in O(16) in the obvious way

$$\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & O(6) \end{array}\right) \tag{55}$$

This breaks O(16) down to O(10).

We will not assume the six compact dimensions to be CP^3. The Dirac equation for this system can be analyzed in general. The zero modes are always ordinary families, never antifamilies and never other representations of O(10). The families are neutral under any continuous symmetries of the six dimensional space B.

The number of families always equals the Euler characteristic of B. Here are some examples:

S^6	2	families	
$S^2 \times S^4$	4	families	
$S^2 \times S^2 \times S^2$	8	families	(56)
CP^3	4	families	
$S^3 \times S^3$	0	families	

In ten dimensions, the number of families is always even. A similar model in eight or twelve dimensions can give an odd number of families.

Why is the number of families equal to the Euler characteristic of B? With $A_\mu{}^i{}_j \sim \omega_\mu{}^i{}_j$, the O(16) spinor index behaves as an extra Lorentz spinor index of B. The fermi field is therefore a spinor-spinor, a field with two independent six dimensional spinor indices. Such a field is equivalent to the de Rham complex of antisymmetric tensor fields. The Dirac operator becomes the d and d^* operators on differential forms, and the number of zero modes (weighted by chirality) is the Euler characteristic.

VII. OTHER CONTEXTS

The preceding discussion has made clear that in Riemannian geometry it will be very difficult to obtain realistic fermion quantum numbers as zero modes of physically acceptable wave operators. One way out, considered in the last section, is to introduce elementary gauge fields. In the context of

Kaluza-Klein theory, this is a rather disappointing possibility. What other alternatives might there be?

We must modify Riemannian geometry in some way. One possibility is that the tangent space group in 4+n dimensions is not O(1, 3+n) but a smaller group G. (This possibility has been considered by Davidson[30] and in much detail by Weinberg.[31]) The smaller group will have more representations, in general; some of the new representations may correspond to new options for the spins of massless particles.

For instance, G may be O(1,3) × O(n). This group has a representation with spin 1/2 under O(1,3) and (say) spin 5/2 under O(n). Since the "true" spin is 1/2, there are no timelike modes with wrong metric and no difficulty in writing a sensible wave operator. In fact, in the ground state the spin (1/2,5/2) field may be regarded as a tensor spinor $\psi_{ij\alpha}$ with i and j constrained to be tangent to the Kaluza-Klein dimensions; a satisfactory wave equation is $\not{D}\psi_{ij} = 0$. (Once one considers fluctuations away from the ground state these formulas do not have a simple generalization, but some generalization would emerge in any theory with restricted tangent space group.) Since the spin 5/2 and higher spin operators in the internal spae can readily have complex zero modes (note the discussion of CP^3 in section (5)), this would enable us to obtain non-vector-like theories after compactification.

If n is even, n=2k, one could consider a theory with tangent space group O(1,3) × U(k). The U(k) group of the internal space is the tangent space group in Kahler geometry. In Kahler geometry, one may write many variants of the Dirac equation (corresponding to the many representations of U(k) = SU(k) × U(1) which do not extend to representations of O(2k)) which lack analogues in Riemannian geometry. This is the subject of $\bar{\partial}$ cohomology, a major subject in Kahler geometry. The modified Dirac equations of Kahler geometry can readily have complex zero modes. For instance, the equations for a charged spin 1/2 field interacting with a magnetic monopole on S^2 or CP^3 can be regarded as equations in Kahler geometry; therefore, the model of section three and the first two models of section six can be viewed in this light.

Of course, it is disappointing to consider tangent space groups like $O(1,3) \times O(n)$ or $O(1,3) \times U(k)$ that are product groups. One would much prefer a unified group, even if smaller than $O(1, 3+n)$. However, Weinberg[31] has shown that if one desires Lorentz invariance in four dimensions producct groups are the only possibilities.

If one is willing to envisage a product group G, one must still find a sensible equation -- replacing the Einstein equation -- for the time dependence of the G connection. And presumably one must face at some point the unrenormalizability of quantum field theory in 4+n dimensins, which is likely to persist in this context.

A more drastic modification of Riemannian geometry would be to assume that the underlying theory is not a field theory of the usual kind but a theory of some other type. For instance, at present the supersymmetric string theories in ten dimensions[7] would appear to be very attractive candidates -- especially the n=2 theory which is chirally asymmetric and anomaly free. This theory naively reduces at low energies to ten dimensional supergravity, but unlike that theory,[7,32] it is likely to be a finite theory to all orders.

Naive compactification of the string theory proceeds via ten dimensional field theory and suffers from the problems of ten dimensional field theory in describing fermion quantum numbers. However, there may be "inherently stringy" ways to compactify the string theory directly to four dimensions, without ten dimensional field theory as an intermediate stage. The rules and problems of Riemannian geometry might not apply in such a case -- though I hope some of the concepts of this paper would be relevant. I would consider this the most attractive possibility, but unfortunately with the present incomplete understanding of the string theory, it is difficult to pursue this possibility.

I will, however, discuss one aspect of the problem. The string theory has a single dimensionless coupling constant λ. (It is essentially Newton's constant written in units of the Regge slope.) It is generally believed that λ is an arbitrarily adjustable constant. If so, the fermion quantum numbers cannot depend on λ. I believe that instead, when the string theory is more fully understood, it will be seen that (with proper normalization of λ) the mathematical consistency of the theory will require that $1/\lambda$ be an integer. The action of the string field theory which has been partly constructed[33] is rather analogous to the effective action of the large N expansion in QCD. We now know [34] that the large N effective action is multi-valued, defined only modulo $2\pi N$. This means that, internally to the $1/N$ expansion, N must be an integer; this is analogous to the quantization of coupling constants in some 2+1 dimensional field theories.[35] The close analogy between λ and $1/N$ strongly suggests that the string field theory action is likewise multi-valued, defined only modulo $2\pi/\lambda$. This would mean that $1/\lambda$ would have to be an integer n; the proper quark and lepton quantum numbers might emerge only for a definite value of n.

VIII. MASSLESS SCALARS

The main focus of this paper is the question of obtaining massless fermions as zero modes in Kaluza-Klein theory. However, massless bosons are also important. Massless spin one and spin two bosons have a well-known origin in Kaluza-Klein theory, and arise for a simple reason, as reflections of unbroken local symmetries. Massless scalars do not have such a simple rationale. Yet a good explanation of the existence of massless charged scalars would be of utmost importance: it would offer a solution[36] to the problem of the existence of widely disparate mass scales in physics.

Any bose field of a Kaluza-Klein theory might have modes that would be seen as charged scalars in four dimensions. We must consider, then the gravitational field; the antisymmetric tensor fields of certain supergravity and other theories; and gauge fields. (More generally, in suitable backgrounds, different fields may mix; mixed modes may be considered under (i) or (iii) below.)

In general terms, we do not want scalars that are massless for reasons of symmetry. The only scalars kept massless by any symmetry argument of the usual sort are Goldstone bosons, which are always neutral under any unbroken gauge symmetries and hence are no help in solving the hierarchy problem. Moreover, for a Goldstone boson any potential at all is forbidden; there is no reason for a Goldstone boson field to acquire tiny but non-zero vacuum expectation values. We wish a more subtle argument for bosonic zero modes, perhaps a topological argument, which will forbid mass terms but allow quartic self-couplings.

Let us consider in turn the case of gravitational, antisymmetric tensor, and Yang-Mills zero modes.

(i) It seems that very little is known about the conditions under which some oscillations in the geometry of a compact space B will correspond to massless scalars. If B is Ricci flat, $R_{\mu\nu} = 0$, a "breathing mode" in which the geometry of B is uniformly dilated corresponds to a massless scalar, because the equation $R_{\mu\nu} = 0$ has a scaling symmetry and does not determine the radius of B. This mode is always neutral under continuous symmetries, so is not helpful in solving the hirarchy problem. The scaling symmetry of the classical equation $R_{\mu\nu} = 0$ is not a symmetry of quantum Kaluza-Klein theories, and therefore[37] (unless there is an unbroken supersymmetry) these modes get non-zero mass at the one loop level.

In models with an unbroken supersymmetry at the tree level, there are some known cases in which some oscillations in the metric of B correspond to charged massless scalars at the tree level.* Little is known about the possibility of eventual supersymmetry breaking in these models.

(ii) Many supergravity theories contain antisymmetric tensor fields, for instance the third rank antisymmetric tensor field A_{ijk} of eleven dimensional supergravity. The Lagrangian is constructed from the curl of A, $F_{ijk\ell} = \partial_i A_{jk\ell} \pm$ cyclic permutations. This curl is invariant under the gauge transformation $A_{ijk} \rightarrow A_{ijk} + (\partial_i \Lambda_{jk} +$ cyclic permutations$)$. The Lagrangian for a k^{th} rank antisymmetric tensor gauge field is

$$\mathcal{L} = \frac{1}{2(k+1)!} \int d^n x \, \left(F_{i_1 \ldots i_{k+1}}\right)^2 \tag{57}$$

The field equations derived from this Lagrangian may readily have zero modes for topological reasons.

---

*I thank M. Duff for a discussion of this point.

The physical interpretation of these modes depends on how many indices of $A_{i_1 \dots i_k}$ are tangent to B and how many are tangent to the space-time directions. If all indices are tangent to B, we get a massless scalar in four dimensions. If all indices but one are tangent to B, we get a massless spin one particle in four dimensions. If all indices but two are tangent to B, we get a massless antisymmetric tensor in four dimensions which again describes a massless scalar. Other cases do not give rise to propagating modes in four dimensions.

In general, the number of zero modes (modulo gauge transformations) of $A_{i_1 \dots i_k}$ with q indices tangent to B is equal to a topological invariant known as the qth Betti number of B. As an example, one may consider in eleven dimensions the spaces M^{pqr} with SU(3)×SU(2)×U(1) symmetry.[4] For most of these spaces, the first and third Betti numbers vanish, so one does not get massless scalars. (The exceptions are $CP^2 \times S^3$ and $CP^2 \times S^3/Z^k$, for which the third Betti number is one and one gets one massless scalar in four dimensions; and $CP^2 \times S^2 \times S^1$, for which the first and third Betti numbers are one and one gets two massless scalars.) However, for the M^{pqr} the second Betti number is one (except for $CP^2 \times S^2 \times S^1$, where it is two), so one would get one (or two) massless spin one particles in four dimensions (in addition to the gauge fields of SU(3)×SU(2)×U(1) coming from the metric tensor). These massless spin one particles do not have minimal couplings to any matter fields. They interact through derivative couplings, such as magnetic moment couplings to fermi fields. They would give rise to long range spin-spin forces of roughly gravitational strength; presumably this is far too weak to be detectable.

There is, however, an old theorem that zero modes of antisymmetric tensor fields are always neutral under any continuous symmetries. Essentially, this is true because, by the deRham-Hodge theory, zero modes of the q^{th} antisymmetric tensor field on B correspond to topological classes of closed q dimensional submanifolds of B. A continuous symmetry cannot change the topological class of a submanifold, so it leaves invariant all of the zero modes of antisymmetric tensor fields.

Here is an analytical proof.* Let d be the curl operator, so the curl of the antisymmetric tensor field A will be denoted as F = dA. (Thus, $(dA)_{i_1 \cdots i_{k+1}} = (\partial_{i_1} A_{i_2 \cdots i_{k+1}} \pm \text{cyclic permutations})$.) The change in A under a gauge transformation is $A \rightarrow A + d\Lambda$, where Λ is an antisymmetric tensor field with one less index than A. What is a massless mode of the A field? Setting the momentum in the Minkowski directions to zero, we calculate the energy (per unit volume) of an antisymmetric tensor field A by integrating over the compact dimensions. From (57), the integral is $\int_B d\phi\, (dA)^2$, so a massless mode is an antisymmetric tensor field defined on B such that dA = 0. Actually, we want zero modes thta cannot be gauged away, so we want solutions of dA = 0 modulo gauge transformations $A \rightarrow A + d\Lambda$. (Since $d^2 = 0$, any pure gauge $A = d\Lambda$ obeys dA = 0.)

Now let $K^i(\phi^j)$ be an arbitrary Killing vector field, generating the infinitessimal symmetry transformation $\phi^i \rightarrow \phi^i + \epsilon K^i(\phi^j)$. To show that the zero modes of A are neutral under arbitrary continuous symmetries, wemust show that the transformation generated by K^i leaves A unchanged, or more exactly that it leaves A unchanged up to a gauge transformation.

---

*See, for instance, reference (38) for further material on antisymmetric tensor fields.

The generator of the symmetry is known as the Lie derivative. Acting on antisymmetric tensor fields, it takes a particularly simple form and can be conveniently defined as follows. Let i_K be the operation of contraction with K; explicitly, for any antisymmetric tensor field $A_{i_1 \ldots i_k}$, $i_K A$ is an antisymmetric tensor field with one index less: $(i_K A)_{i_1 \ldots i_{k-1}} = K^{i_o} A_{i_o i_1 \ldots i_{k-1}}$. Then $\mathcal{L}_K$, on differential forms, can be defined as follows:

$$\mathcal{L}_K = d\, i_K + i_K\, d \tag{58}$$

Thus, the infinitessimal change of A under the transformation generated by K is $A \to A + \epsilon \mathcal{L}_K A$. Equation (58), which may appear unfamiliar, can be seen to agree with the formulas in books on general relativity.

What we must show is that acting on any zero mode, $\mathcal{L}_K$ vanishes up to a gauge transformation. This is not difficult. A zero mode A obeys $dA = 0$, so $\mathcal{L}_K A = (d i_K + i_K d)\, A = d(i_K A)$, but $d(i_K A)$ is the change in A under a gauge transformation (with gauge parameter $\Lambda = i_K A$).

The Atiyah-Hirzebruch theorem, which is more delicate, may be seen as a generalization of this classical result to the spin 1/2 case.

(iii) Now we consider zero modes of gauge fields. Abelian gauge fields are a special case of the antisymmetric tensor fields just considered, but elementary non-abelian gauge fields A_i^a that might be present in the $4+n$ dimensional theory raise different issues.

Zero modes of A_i^a in which the vector index i is tangent to the compact dimensions will be observed as massless scalars in four dimensions. It is certainly possible -- classically -- to obtain charged massless scalars in this way. The simplest example is the original Kaluza-Klein theory, the sole extra dimension being a circle S^1. If elementary non-abelian gauge fields are posited, then the mode in which $A_5{}^a$ is a constant, independent of position on the circle, gives rise in four dimensions to a multiplet of massless scalars in the adjoint representation.

In this case, there are (classically) no quartic interactions, but that is not a general property. In 4+n dimensions, if the compact dimensions are circles $S^1 \times \ldots \times S^1$, then the constant modes of all $A_i{}^a$, $i = 5 \ldots 4+n$ are massless scalars, and one obtains at the tree level the quartic potential

$$V(A_i{}^a) = \frac{e^2}{4} \sum_{i,j=5\ldots4+n} \mathrm{Tr}\,[A_i, A_j]^2 \qquad (59)$$

There are other, more elaborate scenarios that lead at the classical level to charged massless scalars as zero modes of Yang-Mills fields. Here is one such scenario. Consider an eight dimensional world in which the four extra dimensions are $S^2 \times S^2$, giving rise to an SU(2)×SU(2) symmetry. Let there be an elementary SU(2) gauge field, one of whose components has an expectation value, breaking SU(2) to U(1). The total unbroken gauge group in four dimensions is then SU(2)×SU(2)×U(1). Suppose that the gauge field expectation value on $S^2 \times S^2$ is of the monopole type, with equal monopole number p on each two-sphere (according to Dirac, p is an integer or a half integer). Then if $p > 1$ it can be shown that for topological reasons the Yang-Mills equations have massless zero modes at the tree level; they have U(1) charge ±1 and transform under SU(2)×SU(2) as $(p, p-1) \oplus (p-1, p)$. Although these modes are massless, they have at the tree level a non-trivial quartic potential.

What these examples have in common is that in all examples I am aware of, there is no reason for the massless modes to remain massless when loop corrections are considered. Indeed, in the first example considered (the massless mode being a constant on the circle), a one loop calculation has been carried out in the abelian case,[37] showing that a nonzero mass does arise unless there is a bose-fermi cancellation.

IX. THE COSMOLOGICAL CONSTANT

Until now we have considered exclusively the zero modes of wave operators. However, the spirit of this paper is to study qualitative problems that might be solved without full understanding of the details of a Kaluza-Klein theory. In that spirit, we will here consider another qualitative problem of outstanding significance: the apparent vanishing in four dimensions of the cosmological constant.

There has been much interest in recent years[39] in the possibility of a dynamical explanation of the vanishing of the cosmological constant -- the possibility of a theory in which regardless of the value of the bare parameters, the cosmological constant spontaneously relaxes to zero. We want, in other words, a mechanism analogous to the axion mechanism for avoiding strong CP violation. Some ideas in this section have been introduced independently in work cited in reference (39).

We would like to find a theory in which the classical equations do not determine the effective cosmological constant -- the actual, macroscopic curvature of four dimensional space. The classical equations should admit for any values of the bare parameters a one parameter family of solutions, depending on an integration constant. The effective cosmological constant should depend on this integration constant. We will look for a mechanism by which the integration constant spontaneously relaxes to the value at which space-time is macroscopically flat.

Of course, it is easy to find a theory in which there is an undetermined integration constant at the classical level. Consider a theory of scalar fields ϕ_i. If the potential energy $V(\phi_i)$ is independent of one of these fields ϕ, the vacuum expectation value of that field will be undetermined at the classical level. However, precisely because $V(\phi^i)$ is independent of ϕ, the effective cosmological constant will be independent of ϕ.

To find a theory with an undetermined integration constant upon which the cosmological constant depends requires a different approach. The only way I know to do this in 3+1 dimensions is to introduce a third rank antisymmetric tensor gauge field $A_{\mu\nu\alpha}$. As in section (8), the Lagrangian is

$$\mathcal{L} = -\frac{1}{48} \int d^4x \, (F_{\mu\nu\alpha\beta})^2 \tag{60}$$

where $F_{\mu\nu\alpha\beta}$ is the gauge invariant curl, $F_{\mu\nu\alpha\beta} = (\partial_\mu A_{\nu\alpha\beta} \pm$ cyclic permutations). If we define the scalar $F = \frac{1}{24} \varepsilon^{\mu\nu\alpha\beta} F_{\mu\nu\alpha\beta}$ ($\varepsilon^{\mu\nu\alpha\beta}$ being the four dimensional Levi-Civita symbol), then the equation of motion from (60) is $\partial_\mu F = 0$. Thus, F is a constant -- but the constant is a constant of integration, not determined by the classical equations. And the cosmological constant definitely depends on F. It equals its value at $F=0$, plus $F^2/8$.

However, this example is too trivial. The equation of motion $\partial_\mu F = 0$ -- which is an exact statement, even quantum mechanically -- appears to tell us that the integration constant F cannot possibly relax to the value at which the effective cosmological constant would vanish.

A less trivial example of the same kind arises if the third rank antisymmetric tensor field is considered in 4+n dimensions. We consider the Lagrangian

$$\mathcal{L} = \int d^{4+n}x \, \sqrt{g} \left(\frac{1}{16\pi G} R - \frac{1}{48} (F_{\mu\nu\alpha\beta})^2 - \Lambda_0\right) \tag{61}$$

For $n=7$, this differs from the bosonic part of 11 dimensional supergravity by the omission of an FFA term (its inclusion would not affect our discussion) and the inclusion of a non-zero bare cosmological constant Λ_0 (forbidden by supersymmetry in eleven dimensions but needed in our discussion to get a solution in which four dimensions are flat). The solution we will discuss is the Freund-Rubin solution,[40] generalized to $\Lambda_0 \neq 0$.

We look for a solution of the classical equations derived from (61) in which space-time takes the form $D^4(\lambda) \times S^7(R)$, where $D^4(\lambda)$ is a four dimensinal de Sitter space of positive, negative, or zero curvature λ, and $S^7(R)$ is a seven sphere of radius R.

In looking for such a solution, we encounter -- as Freund and Rubin did -- the possibility of a non-zero value of $F = F_{0123}$. As in the four dimensional case, the equations determine only the derivatives of F, not F itself. One may assume an arbitrary value of F, and use the classical equations to solve for λ and R in terms of F. For a whole continuous range of the bare parameters in (61), there is a value of F at which λ -- the curvature of ordinary space -- vanishes.

The gain in going from four to 4+n dimensions is that in 4+n dimensions $F_{\mu\nu\alpha\beta}$ has a non-trivial dynamics with propagating modes as well as an integration constant. One can at least imagine that there may be a quantum mechanical mechanism by which the integration constant F spontaneously relaxes to some special value -- hopefully the value at which $\lambda=0$. But what might this mechanism be?

In condensed matter physics there are some fascinating systems with the following properties. (For recent theoretical discussions and references to previous work see reference (41).) The macroscopic equations have a one parameter family of solutions depending on an integration constant x. There is a critical value of x, say $x=x_0$, such that the classical solution is stable against small oscillations for $x > x_0$ and unstable for $x < x_0$.

What value of x would be observed physically? One would hardly expect to observe the unstable solutions of $x < x_0$, but one might expect that, depending on initial conditions, any stable solution with $x > x_0$ would be accessible.

The surprise is that it is claimed[41] that a whole class of systems spontaneously relaxes -- by means that are not well understood -- to the threshold of stability, $x = x_0$. The fact that the mechanism is so little understood in the condensed matter context invites the speculation that a similar phenomenon could occur in the case of the cosmological constant. Although $\lambda=0$ (flat space) is not exactly a threshold of stability in any obvious sense, it is certainly the dividing point between two qualitatively different regimes, de Sitter space and anti de Sitter space. Anti de Sitter space has a positive energy theorem, which de Sitter space does not;[42] but de Sitter space has a global initial value hypersurface, which anti de Sitter space does not.[43] They are certainly very different. Perhaps the little-understood mechanism by which the condensed matter systems relax has a "cosmological" analogue -- though it is not yet clear whether it is de Sitter or anti de Siter space that should correspond in the analogy to $x > x_0$.

I would like to thank R. Stong for useful dicussions and for drawing my attention to references (44) and (45), and to thank W. Browder and especially W. Hsiang for explaining various aspects of cobordism theory.

References

1. Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys. Kl (1921) 966; O.Klein, Z. Phys. 37 (1926) 895, Arkiv. Mat. Astron. Fys. B. 34A (1946), Helv. Phys. Acta Suppl. IV (1956) 58; A. Einstein and P. Bergmann, Ann. Math. 39 (1938) 683; B. De Witt, in Relativity, Groups, and Topology, ed. B. De Witt and C. De Witt (New York, Gordon and Breach, 1964), and Dynamical Theory of Groups and Fields (New York, Gordon and Breach, 1965); J. Rayski, Acta Phys. Pol. 27 (1965) 89; R. Kerner, Ann. Inst. H. Poincaré, 9 (1968) 143; A. Trautman, Rep. Math. Phys. 1 (1970) 29; Y.M. Cho, J. Math. Phys. 16 (1975) 2029; Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711; L.N. Chang, K.I. Macrae, and F. Mansouri, Phys. Rev. D13 (1976) 235.
2. J. Scherk and J.H. Schwarz, Phys. Lett. 57B (1975) 463; E. Cremmer and J. Scherk, Nucl. Phys. B103 (1976) 393, B108 (1976) 409.
3. A complete list is impossible, but here are some of the recent papers: Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, Nucl. Phys. B122 (1977) 57; J.-F. Luciani, Nucl. Phys. B135 (1978) 111; M.J. Duff and D. Toms, in Unification of Fundamental Interactions II, ed. S. Ferrara and J. Taylor (Cambridge University Press, 1982); M. Duff, in Supergravity 81, ed. S. Ferrara and J. Taylor (Cambridge University Press, 1982); M.J. Duff and C.N. Pope, Imperial College preprint ICTP/82-83/7 (1983); M.J. Duff, B. Nilsson and C. Pope, Texas preprint UTTG-5 (1983); F. Englert, M. Rooman, and P. Spindel, Phys. Lett. 127B (1983) 47; C. Orzalesi, Parma preprint IFPR/TH63 (1980); W. Mecklenburg, Phys. Rev. D21 (1980) 2149; A. Salam and J. Strathdee, Annals of Physics 131 (1982) 316; C. Destri, C. Orzalesi, and P. Rossi, NY University preprint NYU/TR8/82 (1982); S. Weinberg, Phys. Lett. 125B (1983) 65; P. Candelas and S. Weinberg, Texas

preprint UTTG-6-83 (1983); M.F. Awada, Phys. Lett. *127B* (1983) 413; F. Mansouri and L. Witten, Phys. Lett. *127B* (1983) 341.

4. E. Witten, Nucl. Phys. *B186* (1981) 412.
5. L. Castellani, R. D'Auria, and P. Fré, Torino preprint (1983).
6. For reviews, see P. Von Nieuwenhiuzen, Phys. Rep. *68* (1981) 191; J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton University Press, 1983; J. Gates, M. Grisaru, M. Rocek, and W. Siegel, *Superspace* (Addison-Wesley, 1983).
7. M.B. Green and J.H. Schwarz, Nucl. Phys. *B181* (1981) 502, *B198* (1982) 252, 441; M.B. Green, J.H. Schwarz, and L. Brink, Nucl. Phys. *B198* (1982) 474; J.H. Schwarz, Phys. Rep. *89* (1982) 223.
8. This concept and some others discussed in this section are largely due to H. Georgi; see, for instance, Nucl. Phys. *B156* (1979) 126.
9. M. Gell-Mann, P. Ramond, and R. Slansky, in *Supergravity* ed. P. Van Nieuwenhuizen and D.Z. Freedman (North-Holland Publishing Company, 1979); F. Wilczek and A. Zee, Princeton Report, 1979 (unpublished), Phys. Rev. *D25* (1982) 553; F. Wilczek, in *Proceedings of the 1979 International Symposium on Lepton and Photon Interactions at High Energies*, ed. T. Kirk and H. Abarbanel (Fermilab, Batavia, Illinois, 1980); H. Georgi and E. Witten (unpublished); H. Sato, Phys. Rev. Lett. *45* (1980) 1997; A. Davidson, K.C. Wali, P.D. Mannheim, Phys. Rev. Lett. *45* (1980) 35; J. Maalampi and K. Enqvist, Phys. Lett. *97B* (1980) 217; R. Cahn and H. Harari, Nucl. Phys. *B176* (1980) 135; R. Mohapatra and B. Sakita, Phys. Rev. *D21* (1980) 1062; R. Gatto and J. Casabuoni, Phys. Lett. *93B* (1980) 46; J.F. Kim, Phys. Rev. Lett. *45* (1980) 1916.
10. M. Gell-Mann, P. Ramond, R. Slansky, reference (9).
11. L. Palla, Proc. 1978 Tokyo Conference on High Energy Physics, p. 629.

12. N.S. Manton, Nucl. Phys. B193 (1981) 391; G. Chapline and N. Manton, Nucl. Phys. B184 (1981) 391; G. Chapline and R. Slansky, Nucl. Phys. B209 (1982) 461.
13. C. Wetterich, Nucl. Phys. B187 (1981) 343, Phys. Lett. 110B (1982) 379, 113B (1982) 377, CERN preprints TH. 3310 (1982), 3488 (1982), 3517 (1983).
14. S. Randjbar-Daemi, A. Salam, and A. Strathdee, Nucl. Phys. B214 (1983) 491, Phys. Lett. 124B (1983) 349, Trieste preprint (1983).
15. M. Atiyah and F. Hirzebruch, in Eassays on Topology and Related Topics, ed. A. Haefliger and R. Narasimhan (Springer Verlag, 1970), p. 18.
16. E. Witten, J. Diff. Geom. 17 (1982) 661.
17. W. Nahm, Nucl. Phys. B135 (1978) 149.
18. L. Alvarez-Gaumé and E. Witten, Harvard preprint (1983).
19. L. Brink, J.H. Schwarz, and J. Schwarz, Nucl. Phys. B121 (1977) 77.
20. M.F. Atiyah, R. Bott, and A. Shapiro, Topology 3 (1964) Suppl. 1, 3.
21. A. Lichnerowicz, C.R. Acad. Sci. Paris, Serie A-B 257 (1963) 7.
22. H.B. Lawson, Jr. and S.-T. Yau, Comm. Math. Helv. 49 (1974) 232.
23. M.T. Grisaru, H.N. Pendleton, and P. Van Nieuwenhuysen, Phys. Rev. D15 (1977) 996; F.A. Berends, J.W. van Holten, B. de Wit, and P. van Nieuwenhuysen, Nucl. Phys. B154 (1979) 261, Phys. Lett. 83B (1979) 188, J. Phys. A13 (1980) 1643; C. Aragone and S. Deser, Phys. Lett. 85B (1979) 161, Phys. Rev. D21 (1980) 352; K. Johnson and E.C.G. Sudarshan, Ann. Phys. 13 (1961) 126; G. Velo and D. Zwanziger, Phys. Rev. 186 (1967) 1337.
24. R. Bott, in Differential and Combinatorial Topology (Princeton University Press, 1965).
25. See, for example, S. Coleman, in The Unity of the Fundamental Interactions, ed. A. Zichichi (Plenum Press, 1983).

26. M.F. Atiyah and I.M. Singer, Ann. Math. 87 (1968) 484, 546; M.F. Atiyah and G.B. Segal, Ann. Math. 87 (1968) 531; M.F. Atiyah and R. Bott, Ann. Math. 87 (1968) 456.
27. L.D. Landau and E.M. Lifschitz, *Quantum Mechanics*, trans. J.B. Sykes, J.S. Bell (Oxford, Pergamon Press, 1965) p. 371.
28. A. Borel and F. Hirzebruch, Amer. J. Math., 80 (1958) 458, 81 (1959) 315.
29. For an exposition of the topological obstruction to defining spinors, see, for example, S.W. Hawking and C.N. Pope, Phys. Lett. 73B (1978) 42.
30. A. Davidson, Ph.D. thesis (Technion, Israel),and private communication.
31. S. Weinberg, to be published.
32. J.H. Schwarz, Phys. Rept. 89 (1982) 223, and references therein; C.H. Ragadiakos and J. Taylor, Phys. Lett. 124B (1983) 201.
33. M. Green and J.H. Schwarz, Cal Tech preprints, 1982-3.
34. E. Witten, Nucl. Phys. B223 (1983) 422.
35. S. Deser, R. Jackiw, and S. Templeton, Phys. Rev. Lett. 48 (1982) 975, Ann. Phys. (NY) 140 (1982) 372.
36. S. Weinberg, Phys Lett. 62B (1976) 111; S. Weinberg and E. Gildener, Phys. Rev. D13 (1976) 333.
37. T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141; D. Pollard, ICTP preprint 82-83/11; M.A. Rubin and B.D. Roth, Phys. Lett. 127B (1983) 55; T. Appelquist, A. Chodos and E. Meyers, Phys. Lett. 127B (1983) 51; T. Appelquist and A. Chodos, Yale preprint YTP 83-05 (1983); Y. Hosotani, U. Pennsylvania preprint (1983); M.A. Awada and D.J. Toms, Imperial College preprints (1983).
38. R. Bott, L.W. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag (1982).

39. S.W. Hawking, Nucl. Phys. B144 (1978) 349; in Unified Theories of Elementary Particles (Springer-Verlag, 1982); talk at Shelter Island Conference (June, 1983); L.F. Abbott, Brandeis University preprint (1981); V.A. Rubakov and M.E. Schaposnikov, Phys. Lett. 125B (1983) 139.

40. P.G.O. Freund and M.A. Rubin, Phys. Lett. B97 (1980) 223; A. Aurelia, H. Nikolai, and P.K. Townsand, Nucl. Phys. B176 (1980) 509.

41. G. Dee and J.S. Langer, Phys. Rev. Lett. 50 (1983) 383; R.C. Brower, D.A. Kessler, J. Koplik, and H. Levine, Phys. Rev. Lett. 51 (1983) 1111.

42. L. Abbott and S. Deser, Nucl. Phys. B195 (1982) 76; P. Breitenlohner and D.Z. Freedman, Ann. Phys. (NY) 144 (1982); G.W. Gibbons, C.M. Hull, and N.P. Warner, Nucl. Phys. B218 (1983) 173.

43. S.W. Hawking and G.F.R. Ellis, The Large Scale Structure of Space-Time (Cambridge University Press, 1973).

44. C. Kosniowski, Actions of Finite Abelian Groups (Pitman, 1978).

45. O.R. Musin, Math. USSR Sbornik 44 (1983) 325.

DIMENSIONAL REDUCTION OF FERMIONS IN GENERALIZED GRAVITY*

C. WETTERICH

Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

Received 21 February 1984

We discuss possibilities of obtaining chiral four-dimensional fermions from dimensional reduction of pure higher dimensional gravity. We explore a modification of riemannian geometry where the Lorentz rotations are treated in close analogy to usual gauge theories. The metric is not the product of two vielbeins and the vielbein may not be invertible everywhere. The bundle structure of Lorentz transformations is distinguished from the bundle structure of tangent space rotations and the gravitational index theorems have to be modified for this case. We also investigate noncompact internal spaces with finite volume in the context of riemannian geometry. Chiral fermions are obtained in the latter case.

As a byproduct of this work, we find that for the usual torsion theories the Dirac operator is not the relevant mass operator for dimensional reduction of fermions.

1. Introduction

Gravity in more than four space-time dimensions can give a unified description of gauge interactions and four-dimensional gravity if all spacelike dimensions except three are curled up to form an "internal" space with very small characteristic length scale. Typically, this length scale is roughly of the order of the Planck length. Isometries of the internal space correspond to gauge symmetries in the effective four-dimensional theory obtained by integrating the action over the internal space.

This idea has been pursued by many authors [1]. There is, however, a major difficulty in constructing a realistic model along these lines: assuming riemannian geometry in higher dimensions and the internal space to be compact, this theory fails to predict chiral fermions in the effective four-dimensional theory. Starting with a spinor coupled to gravity in $d (d > 4)$ dimensions, one always ends in the four-dimensional theory with fermions belonging to vectorlike representations of the gauge group. Typically, all the fermions have masses of the order of the Planck mass, although there are some known cases where for a special choice of parameters the four-dimensional fermions are massless. A natural explanation of the observed small fermion masses requires chiral fermions, where the left-handed and right-handed

---

* Work in part supported by Schweizerischer Nationalfonds.

Weyl spinors belong to inequivalent representations of the low-energy gauge group $SU(3)\times SU(2)\times U(1)$.

Fermion chirality is a very fundamental feature of a theory, and the number of chiral fermions is independent of many details of the theory. In dimensionally reduced gravity theories with compact internal space the number of chiral four-dimensional fermions can be expressed [2] by an index related to the topology of internal space. This index has been discussed in detail by Witten [3], who finds, by applying a theorem due to Atiyah and Hirzebruch [4], that the number of chiral fermions obtained from dimensional reduction of a Weyl spinor coupled to riemannian geometry with $d-4$ spacelike dimensions forming a compact oriented differentiable manifold without boundary is zero. In this case, no chiral fermions can be obtained with respect to any gauge group which corresponds to an isometry of internal space.

A realistic dimensionally reduced gravity theory should circumvent this no-go theorem. One could of course speculate that the observed fermions and gauge bosons are composite rather than fundamental and that the observed gauge symmetries do not correspond to isometries. However, much of the beauty and unification of dimensionally reduced gravity is lost, and very little can actually be calculated beyond the pure speculation. A second alternative is the introduction of supplementary gauge symmetries in higher dimensions in addition to the symmetries of d-dimensional general coordinate and Lorentz transformations. Chiral fermions can be obtained from dimensional reduction in this case [5, 6, 3]. Again, one of the main motivations to study higher dimensional gravity, the unification of gravity and gauge symmetries, is lost.

Remaining with general coordinate and Lorentz transformations in d dimensions (we denote this group by gen_d) as the only local symmetries of the theory, we must either abandon the assumption that the internal space is compact without boundary or we have to modify riemannian geometry. In this paper we first discuss modifications of riemannian geometry consistent with gen_d being the local transformation group of the theory. More drastic modifications, where the group of d-dimensional Lorentz transformations is replaced by some other transformation group, have been investigated by Weinberg [7]. In the last section we discuss noncompact internal spaces.

We believe that four-dimensional gravity is well described by riemannian geometry with an Einstein-Hilbert action linear in the curvature scalar. This is, however, no indication that d-dimensional gravity should be riemannian too. Indeed, gen_d invariance together with a stable ground state, which is locally a direct product of four-dimensional Minkowski space $\mathfrak{M}^4$ and a D-dimensional internal space K^D with small characteristic length scale, is in general sufficient to guarantee the effective four-dimensional gravity to be riemannian with Einstein-Hilbert action. A ground state $\mathfrak{M}^4\times K^D$ implies that the symmetry group governing four-dimensional gravitational interactions is gen_4. Furthermore, after dimensional reduction, all particles appearing in the effective four-dimensional theory have generically masses of the order of the inverse characteristic length of internal space, except those for which there is a special reason why they have to be massless or very light. Gauge symmetry guarantees massless gauge bosons and a nonvanishing chirality index would guarantee massless fermions. There is a big problem of obtaining massless or very light scalars, the gauge hierarchy problem. (For some remarks on this problem in the context of dimensionally reduced gravity see ref. [8].) Finally, gen_4 invariance guarantees a massless graviton. No other gravitational object except the graviton is expected to be massless or light. Kinetic and interaction terms for the graviton will be dominated by the gen_4 invariant with lowest mass dimension. This is the curvature scalar. Thus the effective four-dimensional action for the graviton will be dominated by an Einstein-Hilbert term. Stability of $\mathcal{M}^4 \times \mathrm{K}^D$ requires Newton's constant to be positive and excludes ghosts or tachyons. We end in four dimensions with usual gravity, whereas all other "geometrical" objects correspond to different tensor fields whose contribution is suppressed by their huge mass of the order of the Planck mass. Since by these reasons we cannot find direct experimental hints on the structure of d-dimensional gravity, we should investigate if a generalized form of d-dimensional gravity can help to generate chiral fermions.

In sect. 2 we discuss the geometrical objects of a generalized gravity theory with gen_d transformation group. We give up the constraints of riemannian geometry that the metric is the square of the vielbein and that the vielbein is covariantly constant. The spin connection is considered to be a free dynamical field. This generalized gravity theory resembles a gauge theory for the Lorentz group. The curvature tensors constructed from the tangent space connection and from the spin connection are not identical any more. We keep the constraint that the tangent space connection is consistent with metricity.

In sect. 3 we give an example for generalized gravity, spheres with vanishing spin connection in arbitrary dimensions D. For D even, the vielbein cannot be invertible everywhere. In our example it vanishes at the south pole. The isometry group is $\mathrm{SO}(D)$ (instead of $\mathrm{SO}(D+1)$ for spheres in riemannian geometry). We find that the Dirac operator has a non-zero chirality index in this case.

In sect. 4 we carry out the dimensional reduction of spinors for our generalized gravity. The usual Dirac operator is in general not the correct mass operator for fermions any more. It has to be replaced by a modified operator in order to guarantee hermiticity. (The modified mass operator is just the hermitian part of the Dirac operator.) These remarks also apply to dimensional reduction of the usual gravitational torsion theories.

In sect. 5 we discuss in detail our example of spheres with vanishing spin connections for $D = 2$. We find that in contrast to the Dirac operator, there are no normalizable massless spinors with respect to the hermitian mass operator.

In sect. 6 we apply index considerations to discuss the prospects of generalized gravity to generate chiral fermions. We find that an invertible vielbein excludes the existence of chiral fermions for compact internal spaces without boundary. However, the no-go theorem described above does not apply if the determinant of the vielbein vanishes at at least one point of the internal manifold.

Finally we discuss in sect. 7 a different alternative to circumvent the no-go theorem: we keep the structure of riemannian geometry, and assume that the internal space is a noncompact space with small finite volume. (Of course, this alternative could be combined with the first one and one may consider generalized gravity with noncompact internal space.) The main idea of unification of gauge interactions with gravity in the context of higher dimensional gravity only requires that the supplementary dimensions form an internal space with a small characteristic length scale and a compact group of isometries. Harmonic expansion and dimensional reduction can be carried out for noncompact internal spaces with finite volume. We give an explicit example for a noncompact space in D dimensions which has finite volume and admits chiral fermions with respect to its isometry group SO(D). For $D=10$, we obtain after dimensional reduction two generations of left-handed fermions belonging to the 16-dimensional complex spinor representation of SO(10). However, many problems remain to be investigated in this approach.

2. Geometrical objects in generalized gravity

We assume that the theory can be formulated on a d-dimensional manifold with one timelike and $d-1$ spacelike dimensions (we use signature $(+,-,-,-\cdots)$) and that this manifold admits an invertible metric $g_{\mu\nu}=g_{\nu\mu}$:

$$g_{\mu\nu}g^{\nu\rho}=\delta_\mu^\rho\,. \tag{1}$$

Under infinitesimal general coordinate transformations

$$x^\mu \to x^\mu+\xi^\mu(x) \tag{2}$$

a covariant vector A_μ in tangent space transforms according to

$$\begin{aligned} &A_\mu \to A_\mu+\delta_\xi A_\mu\,,\\ &\delta_\xi A_\mu=-\partial_\mu\xi^\nu A_\nu-\xi^\nu\partial_\nu A_\mu\,, \end{aligned} \tag{3}$$

with the usual generalizations for other tensors. The derivative $\partial_\mu A_\nu$ is in general not a tensor. In order to describe covariant derivatives transforming as tensors, we have to introduce a connection $\Gamma_{\mu\nu}{}^\lambda$ in tangent space. Under infinitesimal coordinate transformations a connection transforms according to

$$\delta_\xi\Gamma_{\mu\nu}{}^\lambda=-\xi^\rho\partial_\rho\Gamma_{\mu\nu}{}^\lambda+\partial_\rho\xi^\lambda\Gamma_{\mu\nu}{}^\rho-\partial_\mu\xi^\rho\Gamma_{\rho\nu}{}^\lambda-\partial_\nu\xi^\rho\Gamma_{\mu\rho}{}^\lambda-\partial_\mu\partial_\nu\xi^\lambda\,, \tag{4}$$

and the covariant derivative

$$D_\mu A_\nu = \partial_\mu A_\nu - \Gamma_{\mu\nu}{}^\lambda A_\lambda \tag{5}$$

is a tensor. We always will assume that the connection is metric

$$D_\rho g_{\mu\nu} = \partial_\rho g_{\mu\nu} - \Gamma_{\rho\mu}{}^\lambda g_{\lambda\nu} - \Gamma_{\rho\nu}{}^\lambda g_{\mu\lambda} = 0 . \tag{6}$$

Greek indices are raised with the inverse metric $g^{\mu\nu}$ and lowered with $g_{\mu\nu}$ and these operations commute with covariant differentiation:

$$D_\mu A^\nu = \partial_\mu A^\nu + \Gamma_{\mu\lambda}{}^\nu A_\lambda = D_\mu \left(g^{\nu\rho} A_\rho \right) = g^{\nu\rho} D_\mu A_\rho . \tag{7}$$

The commutator of two covariant derivatives defines again a tensor

$$\left[D_\mu , D_\nu \right] A_\rho = - R_{\mu\nu}{}^\lambda{}_\rho A_\lambda - R_{\mu\nu}{}^\lambda D_\lambda A_\rho , \tag{8}$$

where the curvature tensor $R_{\mu\nu}{}^\lambda{}_\rho$ and the antisymmetric part of the connection $R_{\mu\nu}{}^\lambda$ (which, by comparison with (4) is a tensor) are defined by

$$R_{\mu\nu}{}^\lambda{}_\rho = \partial_\mu \Gamma_{\nu\rho}{}^\lambda - \partial_\nu \Gamma_{\mu\rho}{}^\lambda + \Gamma_{\mu\tau}{}^\lambda \Gamma_{\nu\rho}{}^\tau - \Gamma_{\nu\tau}{}^\lambda \Gamma_{\mu\rho}{}^\tau , \tag{9}$$

$$R_{\mu\nu}{}^\lambda = \Gamma_{\mu\nu}{}^\lambda - \Gamma_{\nu\mu}{}^\lambda . \tag{10}$$

The Ricci tensor $R_{\mu\nu}$ and the curvature scalar R are defined by

$$R_{\mu\nu} = R_{\lambda\mu}{}^\lambda{}_\nu , \qquad R = g^{\mu\nu} R_{\mu\nu} . \tag{11}$$

Scalars under general coordinate transformations involving the metric and the connection can be constructed by contractions of the tensors $g_{\mu\nu}$, $R_{\mu\nu}{}^\lambda{}_\rho$ and $R_{\mu\nu}{}^\lambda$, like R, $R_{\mu\nu\sigma\lambda} R^{\mu\nu\sigma\lambda}$, $R_{\mu\nu}{}^\lambda R^{\mu\nu}{}_\lambda$ etc.

We now want to introduce spinors. A spinor is a scalar with respect to general coordinate transformations

$$\delta_\xi \psi = - \xi^\nu \partial_\nu \psi \tag{12}$$

belonging to a fundamental spinor representation of the Lorentz group Spin $(1, d-1)$. A Dirac spinor is a $2^{[d/2]}$ component representation transforming under infinitesimal local Lorentz transformations with parameters $\varepsilon_{mn}(x) = -\varepsilon_{nm}(x)$ according to

$$\delta_{\mathcal{L}} \psi = - \tfrac{1}{2} \varepsilon_{mn} \Sigma^{mn} \psi , \tag{13}$$

where the $d(d-1)/2$ generators Σ^{mn} can be constructed from d Dirac matrices γ^m obeying

$$\{\gamma^m, \gamma^n\} = 2\eta^{mn},$$

$$\Sigma^{mn} = -\tfrac{1}{4}[\gamma^m, \gamma^n], \tag{14}$$

with $\eta^{mn} = \eta_{mn} = \mathrm{diag}(1, -1, -1 \cdots -1)$. Except for $d = 5,\ 7 \bmod 8$, the Dirac spinor is a reducible representation of the Lorentz group and can be decomposed into Majorana, Weyl or Majorana-Weyl spinors (compare ref. [9]). A Lorentz vector e^m transforms according to

$$\delta_{\varepsilon} e^m = \varepsilon^m{}_n e^n, \tag{15}$$

and η^{mn} is an invariant tensor which will be used together with η_{mn} to raise and lower Latin indices. The definition of a covariant spinor derivative requires a spin connection $\omega_{\mu mn} = -\omega_{\mu nm}$ related to the Lorentz group with

$$\delta_{\xi}\omega_{\mu mn} = -\partial_\mu \xi^\nu \omega_{\nu mn} - \xi^\nu \partial_\nu \omega_{\mu mn},$$

$$\delta_{\varepsilon}\omega_{\mu mn} = \varepsilon_m{}^p \omega_{\mu pn} + \varepsilon_n{}^p \omega_{\mu mp} - \partial_\mu \varepsilon_{mn}. \tag{16}$$

Then the covariant derivatives

$$D_\mu \psi = \left(\partial_\mu - \tfrac{1}{2}\omega_{\mu mn}\Sigma^{mn}\right)\psi, \tag{17}$$

$$D_\mu e^m = \partial_\mu e^m + \omega_\mu{}^m{}_n e^n, \tag{18}$$

transform as a spinor and vector respectively under the Lorentz group. The covariant derivative of η_{mn} vanishes. Again the commutator of two covariant derivatives is a tensor

$$\left[D_\mu, D_\nu\right]\psi = -\tfrac{1}{2}F_{\mu\nu mn}\Sigma^{mn}\psi - R_{\mu\nu}{}^\lambda D_\lambda \psi, \tag{19}$$

and $F_{\mu\nu mn}$, the curvature tensor of the spin connection ω, is given by

$$F_{\mu\nu mn} = \partial_\mu \omega_{\nu mn} - \partial_\nu \omega_{\mu mn} + \omega_{\mu m}{}^p \omega_{\nu pn} - \omega_{\nu m}{}^p \omega_{\mu pn}. \tag{20}$$

The construction of a covariant kinetic term for a spinor needs a vielbein $e_\mu{}^m$ transforming as a vector under both general coordinate and Lorentz transformations

$$\delta_\xi e_\mu{}^m = -\partial_\mu \xi^\nu e_\nu{}^m - \xi^\nu \partial_\nu e_\mu{}^m,$$

$$\delta_\varepsilon e_\mu{}^m = \varepsilon^m{}_n e_\mu{}^n. \tag{21}$$

Multiplication with the vielbein converts a Lorentz vector e_n into a covariant vector

$$e_\mu = e_\mu{}^n e_n , \tag{22}$$

and we use the vielbein to convert Latin into Greek indices. If the vielbein has an inverse $e_m{}^\mu$

$$e_m{}^\mu e_\mu{}^n = \delta_m^n , \qquad e_\mu{}^m e_m{}^\nu = \delta_\mu^\nu , \tag{23}$$

(this will not always be the case in our approach), the inverse vielbein can be used to convert Greek into Latin indices

$$e_m = e_m{}^\mu e_\mu . \tag{24}$$

The covariant derivative of the vielbein is

$$D_\mu e_\nu{}^m = \partial_\mu e_\nu{}^m - \Gamma_{\mu\nu}{}^\lambda e_\lambda{}^m + \omega_\mu{}^m{}_n e_\nu{}^n . \tag{25}$$

With the use of the vielbein, a scalar Dirac operator can be constructed:

$$\gamma^\mu D_\mu \psi = g^{\mu\nu} e_{\nu m} \gamma^m D_\mu \psi , \tag{26}$$

and a possible kinetic term for the spinor is

$$\mathfrak{L}_\psi = \tfrac{1}{2} i g^{1/2} \bar{\psi} \gamma^\mu D_\mu \psi + \text{h.c.} \tag{27}$$

with $g^{1/2} = |\det g_{\mu\nu}|^{1/2}$.

A general gen_d invariant gravity theory can be constructed by forming scalars from the various tensors described above. The bosonic geometrical objects of such a general theory are the tensors $g_{\mu\nu}$ and $e_\mu{}^m$ and the connections $\Gamma_{\mu\nu}{}^\lambda$ and $\omega_{\mu m n}$. Riemannian geometry has only one independent geometrical object, the vielbein $e_\mu{}^m$, and the other quantities are related to the vielbein by three constraints:

$$g_{\mu\nu} = e_\mu{}^m e_{\nu m} , \tag{28}$$

$$D_\mu e_{\nu m} = 0 , \tag{29}$$

$$R_{\mu\nu}{}^\lambda = 0 . \tag{30}$$

The first constraint (28) links the metric and the vielbein. It guarantees the existence of an inverse vielbein

$$e_m{}^\mu = g^{\mu\nu} e_\nu{}^n \eta_{nm} = e^\mu{}_m , \tag{31}$$

and implies $e_{\mu m} = e_{m\mu}$. The second constraint (29), together with (28), identifies the bundle structure of the Lorentz group with the principal $\mathrm{SO}(1, d-1)$ bundle of tangent space rotations up to the double covering of $\mathrm{SO}(1, d-1)$ by $\mathrm{Spin}\,(1, d-1)$. From the commutator of two covariant derivatives of the vielbein one obtains

$$0 = \left[D_\mu, D_\nu\right] e_\rho{}^m = F_{\mu\nu}{}^m{}_n e_\rho{}^n - R_{\mu\nu}{}^\lambda{}_\rho e_\lambda{}^m . \tag{32}$$

The curvatures $R_{\mu\nu mn}$ and $F_{\mu\nu mn}$ are identified and all characteristic classes constructed from the connections $\Gamma_{\mu\nu}{}^\lambda$ and $\omega_{\mu mn}$ are identical. According to (25), the spin connection can be expressed through $\Gamma_{\mu\nu}{}^\lambda$ and derivatives of the vielbein. As an example for the consequences of (28), (29), the spin connection cannot vanish identically on a sphere (with euclidean metric) except on S^1, S^3 and S^7, since a vanishing ω requires a trivial bundle structure of the Lorentz group and therefore a trivial tangent bundle. Finally, the third constraint (30) excludes torsion and uniquely determines $\Gamma_{\mu\nu}{}^\lambda$ in terms of derivatives of the metric

$$\Gamma_{\mu\nu}{}^\lambda = \Gamma^{(LC)}{}_{\mu\nu}{}^\lambda \equiv \tfrac{1}{2} g^{\lambda\rho}\left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}\right). \tag{33}$$

The generalized gravity we discuss here abandons the three constraints (28), (29), (30) of riemannian geometry. Eqs. (28) and (29) are now replaced by

$$e_\mu{}^m e_{\nu m} = g_{\mu\nu} + k_{\mu\nu} , \tag{34}$$

$$D_\mu e_{\nu m} = U_{\mu\nu m} , \tag{35}$$

and the constraint (30) may be dropped. The quantities $e_\mu{}^m$, $g_{\mu\nu}$, $\Gamma_{\mu\nu}{}^\lambda$ and $\omega_{\mu mn}$ are then independent quantities and we have new tensors $k_{\mu\nu}$, $U_{\mu\nu m}$, and $R_{\mu\nu}{}^\lambda$. New invariants can be constructed with these tensors. These include the quadratic terms $k_{\mu\nu}k^{\mu\nu}$, $U_{\mu\nu m}U^{\mu\nu m}$ and $R_{\mu\nu}{}^\lambda R^{\mu\nu}{}_\lambda$. For the effective four-dimensional gravity after dimensional reduction, we need the resulting coefficients of these quadratic terms for the four-dimensional tensors to be positive. The ground state is then characterized by

$$k_{\mu\nu}^{(4)} = 0 , \qquad U_{\mu\nu m}^{(4)} = 0 , \qquad R_{\mu\nu}^{(4)\lambda} = 0 . \tag{36}$$

Deviations from four-dimensional riemannian geometry will be suppressed by inverse powers of the Planck mass. In contrast, the effective action for the d-dimensional theory may contain quadratic invariants with negative coefficients, leading to internal spaces with nonvanishing values of $k_{\mu\nu}^{(D)}$, $U_{\mu\nu m}^{(D)}$ or $R_{\mu\nu}^{(D)\lambda}$ and modifying the particle spectrum obtained from dimensional reduction.

The index theorem leading to the no-go theorem for chiral fermions on compact internal spaces is only based on the bundle structure of the transformation groups. It assumes the identification of Lorentz transformations with tangent space rotations up to double covering. We want to circumvent this no-go theorem and are therefore mostly interested in the generalizations (34) and (35). As we will show in sect. 6, the vanishing of $k_{\mu\nu}$ implies a zero chirality index on compact spaces. Our generalization allows us to decouple the bundle structure of Lorentz and tangent space rotations. The Lorentz bundle may be trivial even if the tangent bundle is not. The spin connection $\omega_{\mu mn}$ is completely independent of the tangent space connection $\Gamma_{\mu\nu}{}^{\lambda}$ and we can have vanishing spin connection on an internal manifold forming a sphere in arbitrary dimensions. From the commutator of two covariant derivatives of the vielbein we now obtain

$$\begin{aligned}\left[D_\mu, D_\nu\right] e_\rho{}^m &= F_{\mu\nu}{}^m{}_n e_\rho{}^n - R_{\mu\nu}{}^\lambda{}_\rho e_\lambda{}^m - R_{\mu\nu}{}^\lambda D_\lambda e_\rho{}^m \\ &= D_\mu U_{\nu\rho}{}^m - D_\nu U_{\mu\rho}{}^m = V_{\mu\nu\rho}{}^m, \end{aligned} \tag{37}$$

and for $V_{\mu\nu\rho}{}^m$ nonvanishing the curvatures R and F are different. The characteristic classes constructed from these curvatures can now be different.

The difference between two connections is a tensor. Since the Levi-Civita connection $\Gamma^{(\mathrm{LC})}{}_{\mu\nu}{}^{\lambda}$ defined in (33) transforms as a connection in tangent space, we can always write

$$\Gamma_{\mu\nu}{}^\lambda = \Gamma^{(\mathrm{LC})}{}_{\mu\nu}{}^\lambda + L_{\mu\nu}{}^\lambda, \tag{38}$$

with $L_{\mu\nu}{}^\lambda$ a tensor. Similarly, if an inverse vielbein exists, we can define a connection of the Lorentz group in terms of derivatives of the vielbein and the inverse vielbein:

$$\begin{aligned}\omega(e)_{\mu np} &= e_\mu{}^m \omega(e)_{mnp}, \\ \omega(e)_{mnp} &= -\Omega_{mnp} + \Omega_{npm} - \Omega_{pmn}, \\ \Omega_{mnp} &= -\tfrac{1}{2}\left(e_m{}^\mu e_n{}^\nu - e_n{}^\mu e_m{}^\nu\right) \partial_\mu e_{\nu p}. \end{aligned} \tag{39}$$

One writes

$$\omega_{\mu mn} = \omega(e)_{\mu mn} + K_{\mu mn}, \tag{40}$$

with $K_{\mu mn}$ a tensor. The tensors $L_{\mu\nu}{}^\lambda$ and $K_{\mu mn}$ can be expressed in terms of $R_{\mu\nu}{}^\lambda$ and $U_{\mu\nu m}$:

$$L_{\mu\nu}{}^\lambda = \tfrac{1}{2}\left(R_{\mu\nu}{}^\lambda + R^\lambda{}_{\mu\nu} + R^\lambda{}_{\nu\mu}\right), \tag{41}$$

$$\begin{aligned} K_{mnp} &= e_m{}^\mu K_{\mu np} \\ &= -\tfrac{1}{2}\left\{ R_{mn}{}^\lambda e_{\lambda p} - R_{np}{}^\lambda e_{\lambda m} + R_{pm}{}^\lambda e_{\lambda n} \right. \\ &\qquad \left. + U_{mnp} - U_{nmp} - U_{npm} + U_{pnm} + U_{pmn} - U_{mpn} \right\}. \end{aligned} \tag{42}$$

We note that $e_{\mu m} = e_{m\mu}$ implies $U_{\mu\nu p} = -U_{\mu p\nu}$. For $U_{\mu\nu m} = 0$ one recovers the standard theory of torsion.

3. An example for generalized gravity: spheres with vanishing spin connection

In this section we want to give a concrete example for generalized gravity: spheres with vanishing spin connection in arbitrary dimensions. We will restrict the discussion to the D-dimensional internal space with euclidean signature $\eta_{ab} = -\delta_{ab}$. Once the kinematics of dimensional reduction is carried out [10], the problem of chiral fermions relies entirely on the internal space. For $D = 2 \bmod 4$ the kinematics of dimensional reduction lead to chiral four-dimensional spinors provided the appropriate fermionic mass operator has zero eigenmodes ψ_0 and that the D-dimensional Weyl spinors $\psi_0^+ = \frac{1}{2}(1 + \bar{\Gamma}_D)\psi_0$ and $\psi_0^- = \frac{1}{2}(1 - \bar{\Gamma}_D)\psi_0$ belong to inequivalent representations of the isometry group. ($\bar{\Gamma}_D$ is the D-dimensional generalization of γ^5.)

Consider a space with topology and differentiable structure of a D-dimensional sphere S^D. We use stereographic coordinates. Embedding the sphere with unit radius in a flat $(D+1)$-dimensional space with coordinates z^i, $z^i z^i = 1$, the coordinates y^α on S^D can be induced by stereographic projection

$$y^\alpha = z^\alpha (1 + z^{D+1})^{-1}, \qquad \alpha = 1 \ldots D. \tag{43}$$

This coordinate system can be used everywhere on S^D except on the south pole $z^{D+1} = -1$. Inclusion of the south pole, which requires a second coordinate patch, will be discussed below. To keep the discussion simple, we will assume that the tangent space connection has no torsion: $L_{\alpha\beta}{}^\gamma = R_{\alpha\beta}{}^\gamma = 0$.

The ground state of our theory will be characterized by given values for $e_\alpha{}^a$, $g_{\alpha\beta}$ and $\omega_{\alpha ab}$ which we denote by $\mathring{e}_\alpha{}^a$, $\mathring{g}_{\alpha\beta}$ and $\mathring{\omega}_{\alpha ab}$. (We always assume for the ground state $\mathring{\psi} = 0$.) Isometries of the ground state are those combinations of D-dimensional coordinate and Lorentz transformations which leave $\mathring{e}_\alpha{}^a$, $\mathring{g}_{\alpha\beta}$ and $\mathring{\omega}_{\alpha ab}$ invariant. Such isometries are equivalent with gauge symmetries in the effective four-dimensional action. Let us assume a ground state characterized by

$$\begin{aligned} \mathring{e}_\alpha{}^a &= f(r)\delta_\alpha^a, \\ \mathring{g}_{\alpha\beta} &= -g(r)\delta_{\alpha\beta}, \\ \mathring{\omega}_{\alpha ab} &= 0, \end{aligned} \tag{44}$$

with

$$r^2 = \sum_\alpha y^\alpha y^\alpha. \tag{45}$$

(The standard metric on S^D has $g(r) = 4(1+r^2)^{-2}$.) Obviously, $\mathring{g}_{\alpha\beta}$ is invariant under rotations among the y^α forming the group SO(D). Infinitesimal SO(D) rotations are described by $D(D-1)/2$ Killing vectors $K_z^\alpha(y)$:

$$y^\alpha \to y^\alpha + \delta y^\alpha = y^\alpha + \theta^z K_z{}^\alpha(y), \tag{46}$$

with

$$\theta^z K_z{}^\alpha \equiv \tfrac{1}{2}\theta^{AB} K_{AB}{}^\alpha, \qquad K_{AB}{}^\alpha \equiv -K_{BA}{}^\alpha, \qquad \theta^{AB} = -\theta^{BA}; \qquad A, B = 1 \ldots D,$$
$$K_{AB}{}^\alpha = \left(\delta_A^\alpha \delta_{\beta B} - \delta_B^\alpha \delta_{\beta A}\right) y^\beta. \tag{47}$$

The invariance of the metric is easily checked:

$$\delta \mathring{g}_{\alpha\beta} = -\tfrac{1}{2}\theta^{AB}\left(\partial_\alpha K_{AB}{}^\gamma \mathring{g}_{\gamma\beta} + \partial_\beta K_{AB}{}^\gamma \mathring{g}_{\alpha\gamma} + K_{AB}{}^\gamma \partial_\gamma \mathring{g}_{\alpha\beta}\right) = 0. \tag{48}$$

The vielbein $\mathring{e}_\alpha{}^a$ is also invariant under SO(D) transformations, provided the coordinate transformations (46) are combined with appropriate Lorentz transformations:

$$\delta \mathring{e}_\alpha{}^a = -\tfrac{1}{2}\theta^{AB}\left(\partial_\alpha K_{AB}{}^\gamma \mathring{e}_\gamma{}^a + K_{AB}{}^\gamma \partial_\gamma \mathring{e}_\alpha{}^a\right) - \mathring{e}_\alpha{}^b \mathring{\varepsilon}_b{}^a = 0, \tag{49}$$

with

$$\mathring{\varepsilon}_{ab} = -\tfrac{1}{2}\theta^{AB}(\delta_{aA}\delta_{bB} - \delta_{aB}\delta_{bA}). \tag{50}$$

The isometry group is a linear combination of a special set of coordinate transformations and Lorentz transformations. We note that these Lorentz transformations are independent of y. The inhomogeneous part (see eq. (16)) of the SO(D) transformation of $\mathring{\omega}_{\alpha ab}$ therefore vanishes and $\mathring{\omega}_{\alpha ab} = 0$ is indeed invariant under SO(D) transformations.

The SO(D) gauge transformations act on a spinor as follows [11]:

$$\begin{aligned} \delta\psi &= -\tfrac{1}{2}\theta^{AB} K_{AB}{}^\alpha \partial_\alpha \psi - \tfrac{1}{2}\mathring{\varepsilon}_{ab}\Sigma^{ab}\psi \\ &= -\tfrac{1}{2}\theta^{AB} S_{AB}\psi, \end{aligned} \tag{51}$$

with

$$\begin{aligned} S_{AB} &= K_{AB}{}^\alpha \partial_\alpha - \Sigma_{AB}, \\ \Sigma_{AB} &= \delta_{aA}\delta_{bB}\Sigma^{ab}. \end{aligned} \tag{52}$$

Since $\mathring{\omega}_{\alpha ab}$ vanishes, one immediately finds that a constant spinor χ is a zero eigenmode of the Dirac operator

$$\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}\chi = \mathring{\Gamma}^{\alpha}\partial_{\alpha}\chi = 0\,. \tag{53}$$

(We denote Γ^a the "internal Dirac matrices" $\{\Gamma^a, \Gamma^b\} = 2\eta^{ab}$, $\mathring{\Gamma}^{\alpha} = \mathring{e}^{\alpha}{}_{a}\Gamma^{a}$.) Under SO($D$) transformations, χ transforms

$$\delta\chi = \tfrac{1}{2}\theta^{AB}\Sigma_{AB}\chi\,, \tag{54}$$

and for D even, the Weyl spinors

$$\chi^{+} = \tfrac{1}{2}\left(1 + \bar{\Gamma}_{D}\right)\chi\,, \qquad \chi^{-} = \tfrac{1}{2}\left(1 - \bar{\Gamma}_{D}\right)\chi\,. \tag{55}$$

belong to inequivalent spinor representations of SO(D). For $D = 4 \bmod 4$, χ^{+} and χ^{-} belong to real or pseudoreal representations, whereas for $D = 2 \bmod 4$, χ^{+} and χ^{-} are in complex representations conjugate to each other. For example, for $D = 2$ χ^{+} and χ^{-} have opposite abelian charges. For $D = 10$, χ^{+} belongs to a 16 representation and χ^{-} transforms as $\overline{16}$. However, before asserting the existence of chiral four-dimensional spinors in this model, we have to discuss the kinematics of dimensional reduction for our generalized gravity (see next section) and to analyze a complete harmonic expansion of an arbitrary spinor function $\psi(y)$.

In order to understand the bundle structure of our example, we have to investigate what happens in the coordinate patch covering the south pole. In particular, we will find that the vielbein $\mathring{e}_{\alpha}{}^{a}$ must vanish at the south pole. On the northern hemisphere (NH) we have introduced stereographic coordinates y^{α}. They can be extended everywhere on S^D except on the south pole S. To cover the whole southern hemisphere (SH), we need a different projection and define coordinates

$$y'^{\alpha} = a^{\alpha}{}_{\beta}z^{\beta}(1 - z^{D+1})^{-1}\,, \qquad \alpha, \beta = 1 \ldots D\,, \tag{56}$$

with $a^{\alpha}{}_{\beta}$ to be chosen either $a^{\alpha}{}_{\beta} = \delta^{\alpha}_{\beta}$ or $a^{\alpha}{}_{\beta} = \mathrm{diag}(-1, 1 \ldots 1)$ in order to have the standard orientation. In turn, the coordinates y' cannot be extended to the north pole (N), $z^{D+1} = +1$. The relation between the coordinates y and y' in the overlap of NH and SH is characterized by a transition function $\phi_{\alpha}{}^{\beta}$:

$$\begin{aligned} y'^{\alpha} &= a^{\alpha}_{\beta}y^{\beta}r^{-2}\,, \qquad r^{2} = \sum_{\alpha} y^{\alpha}y^{\alpha} \\ y^{\alpha} &= a^{\alpha}_{\beta}y'^{\beta}r'^{-2}\,, \qquad r'^{2} = \sum_{\alpha} y'^{\alpha}y'^{\alpha} = r^{-2}\,, \end{aligned} \tag{57}$$

$$\phi_{\alpha}{}^{\beta} = \frac{\partial y^{\beta}}{\partial y'^{\alpha}} = a^{\beta}_{\gamma}\left(r'^{-2}\delta^{\gamma}_{\alpha} - 2r'^{-4}y'^{\gamma}y'^{\delta}\delta_{\alpha\delta}\right). \tag{58}$$

On the equator $z^{D+1}=0$, $r^2=r'^2=1$, the transition function describes a rotation. In two dimensions, we may introduce the azimuthal angle φ with

$$\cos\varphi = y'^1/r', \qquad \sin\varphi = y'^2/r'. \tag{59}$$

Using $a^\alpha_\beta = \operatorname{diag}(-1,+1)$, one has

$$\phi_\alpha{}^\beta = \frac{1}{r'^2}\begin{pmatrix} \cos 2\varphi, & -\sin 2\varphi \\ \sin 2\varphi, & \cos 2\varphi \end{pmatrix} \tag{60}$$

This function belongs to a nontrivial homotopy class of $\pi_1(\mathrm{SO}(2))$. In general, the transition functions (58) on the equator are mappings from S^{D-1} on $\mathrm{SO}(D)$ and can be characterized by a homotopy class of $\pi_{D-1}(\mathrm{SO}(D))$.

We next want to describe vector fields in the tangent space of the sphere. We may parametrize such a field by $A_\alpha(y)$ in NH and $A'_\alpha(y')$ in SH. If A and A' are to describe the same field, they must be related in the overlap region of NH and SH by the transition function:

$$A'_\alpha(y') = \phi_\alpha{}^\beta A_\beta(y), \tag{61}$$

for $y'=y'(y)$ according to relation (57). The transition function characterizes the bundle structure related to the tangent space of the sphere. The vielbein $\mathring{e}_\alpha{}^a$ is a set of covariant vectors, but its Lorentz index allows for a more complicated relation between the vielbein $e_\alpha{}^a(y)$ defined in NH and $e'^a_\alpha(y')$ defined in SH. A local Lorentz transformation $\mathcal{L}_b{}^a(y)$

$$\mathring{e}_\alpha{}^a(y) \to \mathring{e}_\alpha{}^b(y)\mathcal{L}_b{}^a(y) \tag{62}$$

leaves the product $\mathring{e}_\alpha{}^a \mathring{e}_{\beta a}$ invariant. Consistency requires

$$\mathring{e}'_\alpha{}^a(y') = \phi_\alpha{}^\beta \mathring{e}_\beta{}^b(y)\mathcal{L}_b{}^a. \tag{63}$$

The transition function for the spinor ψ is governed by the same Lorentz transformation $\mathcal{L}_b{}^a$ as for the vielbein (up to double covering). The Lorentz transformation $\mathcal{L}_b{}^a$ defined in the overlap of NH and SH is again characterized by a homotopy class $\pi_{D-1}(\mathrm{SO}(D))$. They specify the principal bundle $\mathcal{L}(\mathcal{M})$ of Lorentz transformations.

In the standard approach to gravity, the homotopy classes of ϕ^{-1} and $\mathcal{L}$ must be the same. This links the bundle structure of $\mathcal{L}(\mathcal{M})$ to the structure of the tangent bundle $T(\mathcal{M})$. For example, the standard vielbein $\tilde{e}_\alpha{}^a$ induced on S^D from the flat metric in $D+1$ dimensions reads

$$\tilde{e}_\alpha{}^a(y) = \frac{2}{1+r^2}\delta^a_\alpha, \qquad \tilde{e}'_\alpha{}^a(y') = \frac{2}{1+r'^2}\delta^a_\alpha. \tag{64}$$

In the overlap of NH and SH one has

$$\tilde{e}'_{\alpha}{}^{a}=r^{2}\tilde{e}_{\alpha}{}^{a}=\phi_{\alpha}{}^{\beta}\tilde{e}_{\beta}{}^{b}\mathfrak{L}_{b}{}^{a}, \tag{65}$$

implying that $\mathfrak{L}$ and ϕ^{-1} belong to the same homotopy class since their combination must add to a transition function in the trivial homotopy class. For $D=2$ one needs

$$\mathfrak{L}_{b}{}^{a}=\begin{pmatrix} \cos 2\varphi, & \sin 2\varphi \\ -\sin 2\varphi, & \cos 2\varphi \end{pmatrix}. \tag{66}$$

The nontrivial transition function $\mathfrak{L}_{b}{}^{a}$ forbids then the spin connection $\tilde{\omega}_{\alpha ab}$ to vanish everywhere on the sphere: the transition function for $\tilde{\omega}$ involves an inhomogeneous term (compare (16))

$$\tilde{\omega}'_{\alpha ab}(y')=\phi_{\alpha}{}^{\beta}\mathfrak{L}^{-1}{}_{a}{}^{c}\tilde{\omega}_{\beta cd}\mathfrak{L}^{d}{}_{b}+\mathfrak{L}^{-1}{}_{a}{}^{c}\partial_{\alpha}\mathfrak{L}_{cb}, \tag{67}$$

and a vanishing $\tilde{\omega}_{\alpha ab}$ on NH implies a nonvanishing $\tilde{\omega}'_{\alpha ab}$ on SH.

In our generalized approach, $\omega_{\alpha ab}$ is a free field and we can investigate the consequences of $\omega_{\alpha ab}$ vanishing everywhere. In this case $\mathfrak{L}_{b}{}^{a}$ must be a constant Lorentz transformation, independent of y. This transformation belongs to the trivial homotopy class and we will simply take $\mathfrak{L}_{b}{}^{a}=\delta_{b}^{a}$. Starting on NH with a vielbein

$$\mathring{e}_{\alpha}{}^{a}(y)=f(r)\delta_{\alpha}^{a}, \tag{68}$$

the vielbein on SH is given by

$$\mathring{e}'_{\alpha}{}^{a}(y')=\frac{1}{r'^{2}}f\left(\frac{1}{r'}\right)\left(a_{\alpha}^{\beta}\delta_{\beta}^{a}-\frac{2y'^{\gamma}y'^{\delta}}{r'^{2}}\delta_{\delta\alpha}a_{\gamma}^{\beta}\delta_{\beta}^{a}\right). \tag{69}$$

On the two sphere S^2 this vielbein is easily visualized:

$$\mathring{e}'_{\alpha}{}^{a}(y')=\frac{1}{r'^{2}}f\left(\frac{1}{r'}\right)\begin{pmatrix} \cos 2\varphi, & -\sin 2\varphi \\ \sin 2\varphi, & \cos 2\varphi \end{pmatrix}. \tag{70}$$

Going once around the south pole involves a rotation of 4π among the two vectors $\mathring{e}'^{1}_{\alpha}$ and $\mathring{e}'^{2}_{\alpha}$. The vielbein $\mathring{e}_{\alpha}{}^{a}$ is uniquely defined only if it vanishes at the south pole, requiring $f(r)$ to vanish faster than $1/r^2$ for $r\to\infty(\hat{=}\,r'\to 0)$. For D even the vanishing of $\det\mathring{e}_{\alpha}{}^{a}$ at some point on S^D is an intrinsic feature of the trivial bundle structure of the Lorentz transformations. Let us demonstrate this for $D=2$. For trivial $\mathfrak{L}_{b}{}^{a}=\delta_{b}^{a}$, the vielbein $\mathring{e}_{\alpha}{}^{1}$ and $\mathring{e}_{\alpha}{}^{2}$ play the role of normal vector fields with transition functions uniquely determined by ϕ. It is well known that S^2 does not admit vector fields defined everywhere and without zeros. Each e_{α}^{a} must vanish at at least one point on S^2, leading to $\det e_{\alpha}^{a}=0$ at this point. A Lorentz transition function $\mathcal{L}_b{}^a$ belonging to the trivial homotopy class does not change this situation. The reason why in the usual approach to gravity an invertible positive definite metric on S^2 can be constructed from the product of two vielbeins ($g_{\alpha\beta} = e_\alpha{}^a e_{\beta a}$) is related to the nontrivial homotopy class of $\mathcal{L}_b{}^a$ in this case. Conversely, the possibility of chiral fermions in our generalized theory of gravity requires the vanishing of $\mathring{e}_\alpha{}^a$ at S. We will come back to this feature more systematically in sect. 6 and show that zeros of the vielbein at some points on the compact manifold are a necessary condition for the existence of chiral spinors. As an alternative, one may exclude the points where $\mathring{e}_\alpha{}^a$ vanishes from the manifold and consider dimensional reduction on open internal spaces with finite volume. This question is addressed in the last section.

4. Kinematics of dimensional reduction for spinors in generalized gravity

Before proceeding further, we have to generalize the dimensional reduction of fermions [10] to the case when the constraints (28)–(30) of riemannian geometry do not hold. The important modification comes from the fact that the Dirac operator is not hermitian any more. As a consequence, the internal Dirac operator is not the relevant mass operator for the dimensional reduction of fermions. Instead, a modified hermitian operator $\mathcal{D}$ has to be defined ($\mathcal{D}$ is the hermitian part of the Dirac operator). The eigenvalues of this operator restricted to internal space dictate the mass spectrum of fermions after dimensional reduction. These remarks also apply to the usual torsion theories, where zero modes of the Dirac operator have been found for various cases [12].

Let us first discuss the d-dimensional field equations for spinors, following the analysis of ref. [9]. These field equations are equivalent to the complete set of field equations derived from the effective four-dimensional action obtained from dimensional reduction. The Dirac operator can be written in the form

$$\begin{aligned}\gamma^\mu D_\mu &= \gamma^m e^\mu{}_m \left(\partial_\mu - \tfrac{1}{2}\omega_{\mu np}\Sigma^{np} \right)\\ &= \gamma^\mu \partial_\mu + \tfrac{1}{2}\omega^\mu{}_{\mu p}\gamma^p + \tfrac{1}{4}\tilde{\omega}_{mnp}\gamma^{mnp},\end{aligned} \tag{71}$$

with

$$\gamma^{mnp} = \gamma^{[m}\gamma^n\gamma^{p]} \tag{72}$$

the totally antisymmetrized product of three Dirac matrices and $\tilde{\omega}_{mnp}$ totally antisymmetric:

$$\begin{aligned}\tilde{\omega}_{mnp} &= \tilde{\tilde{\omega}}_{[mnp]},\\ \tilde{\tilde{\omega}}_{mnp} &= e^\mu{}_m \omega_{\mu np}.\end{aligned} \tag{73}$$

The hermitian kinetic term (27) for a Dirac or Weyl spinor is of the form

$$\begin{aligned}\tilde{\mathfrak{L}}_\psi &= \tfrac{1}{2} i g^{1/2} \bar{\psi} \gamma^\mu D_\mu \psi + \text{h.c.}\\ &= g^{1/2}\left\{\left(\tfrac{1}{2} i \bar{\psi} \gamma^\mu \partial_\mu \psi + \text{h.c.}\right) + \tfrac{1}{4} i \bar{\psi} \gamma^{mnp} \psi \tilde{\omega}_{mnp}\right\}. \end{aligned} \tag{74}$$

We note that the spinor does not couple directly to the trace of the spin connection $\omega^\mu{}_{\mu p}$ due to the identity

$$i \bar{\psi} \gamma^m \psi + \text{h.c.} = 0. \tag{75}$$

The derivative term can be written

$$\begin{aligned}\tfrac{1}{2} i g^{1/2} \bar{\psi} \gamma^\mu \partial_\mu \psi + \text{h.c.} = i g^{1/2} \bar{\psi} \gamma^\mu \partial_\mu \psi + \tfrac{1}{2} i \partial_\mu \left(g^{1/2} e^\mu{}_m\right) \bar{\psi} \gamma^m \psi\\ - \tfrac{1}{2} i \partial_\mu \left(g^{1/2} \bar{\psi} \gamma^\mu \psi\right). \end{aligned} \tag{76}$$

Using the identities

$$\begin{aligned}\partial_\mu\left(g^{1/2} e^\mu{}_m\right) &= g^{1/2} g^{\mu\nu}\left(\partial_\mu e_{\nu m} - \Gamma^{(\text{LC})}{}_{\mu\nu}{}^\lambda e_\lambda{}^m\right)\\ &= g^{1/2} g^{\mu\nu}\left(D_\mu e_{\nu m} - \omega_{\mu m}{}^p e_{\nu p} + \Gamma_{\mu\nu}{}^\lambda e_{\lambda m} - \Gamma^{(\text{LC})}{}_{\mu\nu}{}^\lambda e_{\lambda m}\right)\\ &= g^{1/2} g^{\mu\nu}\left(U_{\mu\nu m} + L_{\mu\nu}{}^\lambda e_{\lambda m} + \omega_{\mu\nu m}\right)\\ &= g^{1/2}\left(U^\mu{}_{\mu m} + \tilde{L}^\mu{}_{\mu m} + \omega^\mu{}_{\mu m}\right), \end{aligned} \tag{77}$$

we obtain the hermitian part of the Dirac operator up to a total divergence:

$$\begin{aligned}\mathfrak{L}_\psi &= i g^{1/2} \bar{\psi} \mathfrak{D} \psi - \tfrac{1}{2} i \partial_\mu \left(g^{1/2} \bar{\psi} \gamma^\mu \psi\right),\\ \mathfrak{D} &= \gamma^\mu D_\mu + \tfrac{1}{2} U^\mu{}_{\mu m} + \tfrac{1}{2} \tilde{L}^\mu{}_{\mu m}. \end{aligned} \tag{78}$$

The d-dimensional field equations for a Dirac spinor then read

$$\mathfrak{D} \psi = 0. \tag{79}$$

Since $U^\mu{}_{\mu m}$ and $\tilde{L}^\mu{}_{\mu m}$ are tensors, $\mathfrak{D}$ commutes with gen_d transformations as well as $\gamma^\mu D_\mu$. It can be written similar to eq. (71):

$$\mathfrak{D} = \gamma^\mu \partial_\mu + \tfrac{1}{2} \omega_{\text{LC}}(e)^\mu{}_{\mu p} \gamma^p + \tfrac{1}{4} \tilde{\omega}_{mnp} \gamma^{mnp}, \tag{80}$$

where

$$\omega_{\mathrm{LC}}(e)^{\mu}{}_{\mu p} = \omega^{\mu}{}_{\mu p} + U^{\mu}{}_{\mu p} + \tilde{L}^{\mu}{}_{\mu p} \tag{81}$$

has the same transformation properties as $\omega^{\mu}{}_{\mu p}$. It obeys the relation

$$\omega_{\mathrm{LC}}(e)^{\mu}{}_{\mu p} = g^{-1/2}\partial_{\mu}\left(g^{1/2}e^{\mu}{}_{p}\right) \tag{82}$$

and can be expressed entirely in terms of the vielbein, the metric, and their derivatives. Consistent with the observation that the trace of the spin connection does not couple to spinors (75), $\omega^{\mu}{}_{\mu p}$ does not appear in the field equations. All spin connections with the same totally antisymmetric part $\tilde{\omega}_{mnp}$ lead to the same field equations. This is also relevant for the usual torsion theories ($U_{\mu\nu p} = 0$): for vanishing spin connection, constant spinors correspond to zero eigenmodes of $\gamma^{\mu}D_{\mu}$, but in general not to zero eigenmodes of the operator $\mathcal{D}$ which is relevant for the field equations. For $e_{\mu m} = e_{m\mu}$ one has $\omega_{\mathrm{LC}}(e)^{\mu}{}_{\mu p} = \omega(e)^{\mu}{}_{\mu p}$ with $\omega(e)_{mnp}$ defined by eq. (39).

This discussion is easily generalized to spinors obeying a Majorana constraint (Majorana spinors or Majorana-Weyl spinors; for details see ref. [9]). In this case the hermitian kinetic term for spinors is

$$\mathcal{L}_{\psi} = \tfrac{1}{2} i g^{1/2}\left(\tilde{\psi}\gamma^{\mu}\partial_{\mu}\psi + \tfrac{1}{4}\tilde{\psi}\gamma^{mnp}\psi\tilde{\omega}_{mnp}\right). \tag{83}$$

Due to Fermi statistics the bilinear $\tilde{\psi}\gamma^{m}\psi$ vanishes identically. Again, there is no coupling of the spinor to $\omega^{\mu}{}_{\mu p}$. The field equations derived from (83) involve the combination $\partial_{\mu}(g^{1/2}e^{\mu}{}_{m})$ and read

$$\mathcal{D}\psi = 0. \tag{84}$$

Again, the operator $\mathcal{D}$ appears to be the relevant operator. For $d = 2, 3, 4, 8, 9 \bmod 8$ we may decompose a Dirac spinor into two Majorana spinors ψ_1 and ψ_2. If we use the hermitian kinetic term (74) for the Dirac spinor, it decomposes into the sum of two independent kinetic terms for ψ_1 and ψ_2 of the form (83). This would not be true for a kinetic term $ig^{1/2}\bar{\psi}\gamma^{\mu}D_{\mu}\psi$ instead of $ig^{1/2}\bar{\psi}\mathcal{D}\psi$. In this case, there would be nontrivial mixings between ψ_1 and ψ_2. The decoupling of ψ_1 and ψ_2, however, is crucial for a consistent dimensional reduction [10].

Let us next discuss dimensional reduction for a spinor. The treatment will be completely parallel to ref. [10] where the reader is referred to for a more complete and detailed discussion. In view of the discussion of open internal spaces with finite volume in sect. 7 we will only assume that space-time is locally characterized by a direct product of four-dimensional Minkowski space and a D-dimensional internal space with finite volume, characterized by a small length scale. ("Locally" refers to a local region in the usual four coordinates.) It is then useful to express the dependence of fields on the internal coordinates y by a harmonic expansion and to derive an effective four-dimensional action by integration of the action over the internal space. For spinors with a hermitian kinetic term, all information is obtained [10] by expanding a D-dimensional Dirac spinor over the internal space K^D.

We assume that K^D is a D-dimensional space with finite volume (not necessarily compact) and signature $\eta_{ab} = -\delta_{ab}$. We consider an arbitrary normalizable Dirac spinor $\psi(y)$ on K^D—this means that $\int \mathrm{d}^D y\, g_D^{1/2} \psi^\dagger \psi$ is a finite positive number. A normalizable spinor $\psi(y)$ can be seen as a set of $2^{[D/2]}$ normalizable functions. (We remind that $\psi(y)$ are commuting variables in this context.) Since the normalizable functions on K^D form a vector space (of infinite dimension), we can express every normalizable spinor $\psi(y)$ in terms of a normalized basis of this vector space:

$$\psi(y) = a_i \psi_i(y),$$

$$\int \mathrm{d}^D y\, g_D^{1/2} \psi_i^\dagger \psi_j = \delta_{ij},$$

$$g_D = |\det g_{\alpha\beta}|, \tag{85}$$

with finite complex constant coefficients a_i. We want to choose a basis $\{\psi_i(y)\}$ with ψ_i classified by representations of the isometry group and eigenfunctions of an appropriately defined mass operator.

We also assume that K^D admits a group of isometries. For our purpose, this means that *all* geometrical quantities $\mathring{g}_{\alpha\beta}, \mathring{e}_\alpha{}^a, \mathring{\Gamma}_{\alpha\beta}{}^\gamma, \mathring{\omega}_{\alpha bc}$ characterizing the ground state K^D are invariant under some subgroup of gen_D transformations

$$\delta \mathring{g}_{\alpha\beta} = -\theta^z S_z \mathring{g}_{\alpha\beta} = 0 \quad \text{etc}. \tag{86}$$

Here θ^z are infinitesimal transformation parameters independent of the internal coordinates y and the generators S_z are appropriate differential operators fulfilling the usual commutation relations for Lie groups

$$[S_x, S_y] = f_{xyz} S_z \tag{87}$$

(compare the example in sect. 3). The action of the isometry group on a spinor $\psi(y)$ is given by

$$\delta\psi = -\theta^z S_z \psi,$$

$$S_z \psi = K_z{}^\alpha \partial_\alpha \psi - \tfrac{1}{2} \mathring{\varepsilon}_{z,ab} \Sigma^{ab} \psi, \tag{88}$$

with $\mathring{\varepsilon}_{z,ab}$ characterizing the Lorentz transformations needed to guarantee the invariance of the vielbein $\mathring{e}_\alpha{}^a$:

$$\theta^z \mathring{\varepsilon}_{z,ab} = -\mathring{\varepsilon}_{ab} = \mathring{\varepsilon}_{ba},$$

$$\delta \mathring{e}_\alpha{}^a = -\theta^z \left(\partial_\alpha K_z{}^\beta \mathring{e}_\beta{}^a + K_z{}^\beta \partial_\beta \mathring{e}_\alpha{}^a \right) - \mathring{e}_\alpha{}^b \mathring{\varepsilon}_b{}^a = 0. \tag{89}$$

Everywhere where an inverse vielbein $\mathring{e}_a{}^\alpha$ exists, the generators S_z acting on a spinor can be expressed

$$\begin{aligned} S_z\psi = {} & K_z{}^\alpha \mathring{D}_\alpha \psi - \tfrac{1}{2} \mathring{D}_\alpha K_z{}^\beta \mathring{e}_a{}^\alpha \mathring{e}_{\beta b} \Sigma^{ab} \psi \\ & - \tfrac{1}{2} K_z{}^\beta \mathring{U}_{\beta a b} \mathring{e}_a{}^\alpha \Sigma^{ab} \psi + \tfrac{1}{2} K_z{}^\beta \mathring{e}_a{}^\alpha \mathring{e}_{\gamma b} \mathring{R}_{\alpha\beta}{}^\gamma \Sigma^{ab} \psi, \end{aligned} \tag{90}$$

with covariant derivatives $\mathring{D}_\alpha$ constructed from the connections $\mathring{\omega}_{\alpha bc}$ and $\mathring{\Gamma}_{\alpha\beta}{}^\gamma$:

$$\begin{aligned} \mathring{D}_\alpha \psi &= \partial_\alpha \psi - \tfrac{1}{2} \mathring{\omega}_{\alpha ab} \Sigma^{ab} \psi, \\ \mathring{D}_\alpha K_z{}^\beta &= \partial_\alpha K_z{}^\beta + \mathring{\Gamma}_{\alpha\gamma}{}^\beta K_z{}^\gamma, \\ \mathring{U}_{\beta ab} &= \mathring{D}_\beta \mathring{e}_{ab}, \\ \mathring{R}_{\alpha\beta}{}^\gamma &= \mathring{\Gamma}_{\alpha\beta}{}^\gamma - \mathring{\Gamma}_{\beta\alpha}{}^\gamma. \end{aligned} \tag{91}$$

For riemannian geometry ($\mathring{e}_{\alpha a} = \mathring{e}_{a\alpha}$, $\mathring{R}_{\alpha\beta}{}^\gamma = 0$, $\mathring{U}_{\alpha\beta a} = 0$), the generators S_z are called the Lie derivatives of ψ.

Suppose $S_z\psi_i$ can be expanded in terms of the basis functions ψ_j:

$$S_z \psi_i = \psi_j (T_z)_{ji}, \tag{92}$$

$$\left[T_x, T_y \right] = f_{xyz} T_z, \tag{93}$$

with complex constant $(T_z)_{ji}$. The $(T_z)_{ji}$ form an infinite dimensional reducible representation of the isometry group. Reducing to irreducible representations H with components n, we choose a basis ψ_{nHk} (k is a supplementary, so far unspecified index) with

$$\begin{aligned} \psi(y) &= a_{nHk} \psi_{nHk}(y), \\ S_z \psi_{nHk}(y) &= \psi_{n'Hk}(y) (T_z)_{n'n}, \\ \int \mathrm{d}^D y\, g_D^{1/2} \psi_{n'H'k'}{}^\dagger \psi_{nHk} &= \delta_{n'n} \delta_{H'H} \delta_{k'k}. \end{aligned} \tag{94}$$

Next we observe that S_z commutes with the Dirac operator and also with $\mathring{U}^{\alpha}{}_{\alpha a}\Gamma^a$ and $\mathring{\tilde{L}}^{\alpha}{}_{\alpha a}\Gamma^a$:

$$[S_z, \mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}] = 0, \qquad [S_z, \mathring{U}^{\alpha}{}_{\alpha a}\Gamma^a] = 0, \qquad [S_z, \mathring{\tilde{L}}^{\alpha}{}_{\alpha a}\Gamma^a] = 0, \tag{95}$$

implying

$$[S_z, \mathring{\mathfrak{D}}] = 0. \tag{96}$$

If for all (n, H, k) the function $\mathring{\mathfrak{D}}\psi_{nHk}$ can be expanded in the basis $\psi_{n'H'k'}$ we write

$$\mathring{\mathfrak{D}}\psi_{nHk}(y) = \psi_{nHk'}(y) M_{k'k}, \tag{97}$$

with complex constant $M_{k'k}$. $M_{k'k}$ fulfils the rôle of a mass matrix for the four-dimensional spinors after dimensional reduction. This mass matrix is hermitian ($M_{k'k}{}^* = M_{kk'}$) provided the integral over the total divergence $\int \mathrm{d}^D y\, \partial_{\alpha}(g_D^{1/2}\psi_{nHk'}{}^{\dagger}\mathring{\Gamma}^{\alpha}\psi_{nHk})$ vanishes for all ψ_{nHk} (exceptionally, there is no summation over n and H). We then can diagonalize $M_{k'k}$ by unitary transformations and choose a basis of eigenstates of $\mathring{\mathfrak{D}}$:

$$\psi(y) = a_{nHMj}\psi_{nHMj}(y),$$

$$\mathring{\mathfrak{D}}\psi_{nHMj} = M\psi_{nHMj},$$

$$\int \mathrm{d}^D y\, g_D^{1/2}\psi_{n'H'M'j'}{}^{\dagger}\psi_{nHMj} = \delta_{n'n}\delta_{H'H}\delta_{M'M}\delta_{j'j}. \tag{98}$$

The real eigenvalues M correspond to the masses of four-dimensional fermions.

Alternatively, we could also expand in eigenfunctions of $\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}$. However, one now has

$$\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}\psi_{nHk} = \psi_{nHk'}m_{k'k}, \tag{99}$$

where $m_{k'k}$ is *not* hermitian except for riemannian geometry. The mass matrix for four-dimensional fermions after dimensional reduction turns out [10] to be $\frac{1}{2}(m_{kk'} + m^*_{k'k})$. Thus, even if the Dirac operator has zero eigenmodes ($m_{kk'}$ has zero eigenvalues), there are not necessarily massless fermions in the reduced four-dimensional theory. Zero eigenvalues of $m_{kk'}$ do not imply that the hermitian part $\frac{1}{2}(m_{kk'} + m^*_{k'k})$ has zero eigenvalues. We are back to an investigation of the spectrum of the operator $\mathring{\mathfrak{D}}$, which is the only mass operator relevant for dimensional reduction.

5. Spinor harmonics on the sphere for generalized gravity

We now come back to our example of sect. 3 for D-dimensional spheres with vanishing spin connection. We have found zero eigenmodes of the Dirac operator $\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}$ in this case. However, the relevant question is rather the spectrum of $\mathring{\mathcal{D}}$: does $\mathring{\mathcal{D}}$ have zero eigenmodes for this geometry? For $\mathring{e}_{\alpha}{}^{a} = f(r)\delta^{a}_{\alpha}$, $\mathring{g}_{\alpha\beta} = -g(r)\delta_{\alpha\beta}$, $\mathring{\omega}_{\alpha bc} = 0$, $\mathring{R}_{\alpha\beta}{}^{\gamma} = 0$ (eq. (44)) one finds:

$$\mathring{U}^{\alpha}{}_{\alpha a} = \frac{1}{2}\frac{y^{\alpha}}{r}\frac{f}{g}\delta_{\alpha a}\left[(D-2)\frac{g'}{g} + 2\frac{f'}{f}\right],$$

$$\tilde{\mathring{L}}^{\alpha}{}_{\alpha a} = 0, \tag{100}$$

$$\mathring{\mathcal{D}}\psi = \frac{f}{g}\Gamma^{a}\left(\delta^{\alpha}_{a}\partial_{\alpha}\psi + \frac{1}{4}\frac{y^{\alpha}}{r}\delta_{\alpha a}\left[(D-2)\frac{g'}{g} + 2\frac{f'}{f}\right]\psi\right), \tag{101}$$

with $f' = \mathrm{d}f(r)/\mathrm{d}r$, $g' = \mathrm{d}g(r)/\mathrm{d}r$. We remember that any ground state spin connection $\mathring{\omega}_{\alpha ab}$ with vanishing antisymmetric part $\tilde{\mathring{\omega}}_{abc}$ leads to the same operator $\mathring{\mathcal{D}}$.

We first will discuss the harmonic expansion of spinors for the case $D = 2$ in some detail. This gives a simple illustration for the general procedure outlined in the last section, where many of the characteristic features can be seen explicitly. The isometry group is U(1), and we parametrize the orbit of U(1) by a new coordinate φ:

$$y^{1} = r\cos\varphi, \qquad y^{2} = r\sin\varphi. \tag{102}$$

In these coordinates the derivatives read

$$\frac{\partial}{\partial y^{\alpha}} = y^{\beta}\left(\delta_{\alpha\beta}\frac{1}{r}\frac{\partial}{\partial r} - \varepsilon_{\alpha\beta}\frac{1}{r^{2}}\frac{\partial}{\partial\varphi}\right), \tag{103}$$

with $\varepsilon_{12} = -\varepsilon_{21} = 1$, $\varepsilon_{11} = \varepsilon_{22} = 0$. For the Killing vector $\mathrm{K}_{12}{}^{\alpha}$ (eq. 47) one has

$$K_{12}{}^{\alpha}\partial_{\alpha} = -\partial_{\varphi}. \tag{104}$$

The Dirac spinor in two dimensions has two components, which can be decomposed into one-component Weyl spinors ψ^{+} and ψ^{-}:

$$\psi = \begin{pmatrix} \psi^{+} \\ \psi^{-} \end{pmatrix}, \qquad \psi^{\pm} = \tfrac{1}{2}(1 \pm \bar{\Gamma}_{D})\psi. \tag{105}$$

We use the following representation for the Dirac matrices Γ^{a}:

$$\Gamma^{1} = i\tau_{1} = \begin{pmatrix} & i \\ i & \end{pmatrix}, \qquad \Gamma^{2} = i\tau_{2} = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix},$$

$$\bar{\Gamma}_{D} = i\Gamma^{1}\Gamma^{2} = \tau_{3} = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}, \qquad \Sigma^{12} = \tfrac{1}{2}i\tau_{3}. \tag{106}$$

The gauge transformation of a spinor is

$$\delta\psi = -\tfrac{1}{2}\theta^{AB}S_{AB}\psi = \theta\left(\partial_{\varphi} + \tfrac{1}{2}i\tau_3\right)\psi = i\theta Q\psi\,, \tag{107}$$

with $\theta = \theta_{12} = -\theta_{21}$.

We expand an arbitrary normalizable spinor in eigenfunctions of the charge operator Q:

$$\psi(r,\varphi) = a_{nMj}\psi_{nMj}(r,\varphi)\,,$$

$$\psi_{nMj}(r,\varphi) = \frac{1}{\sqrt{2\pi}}\begin{pmatrix} \exp(in\varphi)\chi^{+}_{nMj}(r) \\ \exp(i(n+1)\varphi)\chi^{-}_{nMj}(r) \end{pmatrix}, \tag{108}$$

$$Q\psi_{nMj} = \left(n+\tfrac{1}{2}\right)\psi_{nMj}\,, \tag{109}$$

with n an arbitrary integer*. The functions $\chi^{\pm}_{nMj}(r)$ are normalized

$$\int_0^{\infty} \mathrm{d}r\, r g(r)\left[\left(\chi^{+}_{nM'j'}\right)^{*}\chi^{+}_{nMj} + \left(\chi^{-}_{nM'j'}\right)^{*}\chi^{-}_{nMj}\right] = \delta_{M'M}\delta_{j'j}\,, \tag{110}$$

(without summation over n), implying

$$\int \mathrm{d}^2y\, g_D^{1/2}\left(\psi_{n'M'j'}\right)^{\dagger}\psi_{nMj} = \delta_{n'n}\delta_{M'M}\delta_{j'j}\,. \tag{111}$$

The $\chi^{\pm}_{nMj}(r)$ are a complete basis for normalizable r dependent spinor functions, so that every normalizable spinor depending only on r can be written

$$\psi_n(r) = \begin{pmatrix} \chi^{+}_{n}(r) \\ \chi^{-}_{n}(r) \end{pmatrix} = \sum_{Mj} a_{nMj}\begin{pmatrix} \chi^{+}_{nMj}(r) \\ \chi^{-}_{nMj}(r) \end{pmatrix}. \tag{112}$$

(We note that one may choose a different basis for each n.)

Let us next look for eigenmodes of the mass operator $\mathring{\mathfrak{D}}$:

$$\mathring{\mathfrak{D}}\psi_{nMj} = M\psi_{nMj}\,. \tag{113}$$

One obtains

$$\mathring{\mathfrak{D}}\begin{pmatrix} \exp(in\varphi)\chi^{+}_{nMj}(r) \\ \exp(i(n+1)\varphi)\chi^{-}_{nMj}(r) \end{pmatrix} = \begin{pmatrix} i\dfrac{f}{g}\left(\partial_r + \dfrac{1}{2}\dfrac{f'}{f} + \dfrac{n+1}{r}\right)\exp(in\varphi)\chi^{-}_{nMj}(r) \\ i\dfrac{f}{g}\left(\partial_r + \dfrac{1}{2}\dfrac{f'}{f} - \dfrac{n}{r}\right)\exp(i(n+1)\varphi)\chi^{+}_{nMj}(r) \end{pmatrix}, \tag{114}$$

---

* All representations of U(1) are one-dimensional and labelled by the index n, which corresponds to the index H in the general case of sect. 4. We also note that in contrast to Kaluza's original five-dimensional example there are no neutral fermions in this case, since all fermions carry half-integer charge.

and therefore

$$i\frac{f}{g}\left(\partial_r + \frac{1}{2}\frac{f'}{f} + \frac{n+1}{r}\right)\chi^-_{nMj} = M\chi^+_{nMj},$$

$$i\frac{f}{g}\left(\partial_r + \frac{1}{2}\frac{f'}{f} - \frac{n}{r}\right)\chi^+_{nMj} = M\chi^-_{nMj}. \tag{115}$$

We want to know if there are solutions of (115) with $M = 0$. If we could find a normalizable solution of

$$\left(\partial_r + \frac{1}{2}\frac{f^-}{f} - \frac{n}{r}\right)\chi^+_{n0} = 0, \tag{116}$$

we also would have a zero mode $\chi^-_{n'0}$ with $n' = -(n+1)$. The charges of χ^+_{n0} and $\chi^-_{n'0}$ are $n + \frac{1}{2}$ and $-(n + \frac{1}{2})$ respectively and therefore opposite to each other as required by charge conjugation [10]. A normalizable solution of (116) would indeed induce chiral fermions after dimensional reduction. Eq. (116) is easily integrated and the solution is

$$\chi^+_{n0} = a^+_{n0} r^n f(r)^{-1/2} \tag{117}$$

(with a^+_{n0} an integration constant). Normalizability of this solution requires the integral (110)

$$I_n = \int_0^\infty \mathrm{d}r\, r g(r) |\chi^+_{n0}|^2 \tag{118}$$

to be finite. We have supposed that both $f(r=0)$ and $g(r=0)$ are positive constants and that $f(r)$ and $g(r)$ have no zero for finite r. Taking the limit $r \to 0$, the integral I_n can only be finite if

$$n > -1. \tag{119}$$

Thus n is positive or zero. For $r \to \infty$, consistency requires that $f(r)$ vanishes faster than r^{-2} (compare sect. 3) and for compact spaces one needs

$$\lim_{r\to\infty} g(r) \to r^{-4}. \tag{120}$$

As a consequence, I_n diverges for $r \to \infty$ for all $n \geqslant 0$. There are no normalizable zero modes of $\mathring{\mathscr{D}}$! Coming back to the eigenvalues of the Dirac operator $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$, they are obtained by omitting the term $\frac{1}{2}f'/f$ in eqs. (114)–(116). The solution

$$\tilde{\chi}^+_{n0} = \tilde{a}^+_{n0} r^n \tag{121}$$

is then normalizable for $n=0$. The zero modes discussed in sect. 3 are therefore the only zero modes of the Dirac operator. We observe that in our example of generalized gravity the index [3] is different for the operators $\mathring{\mathcal{D}}$ and $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$. This is due to the fact that both $\mathring{\mathcal{D}}$ and $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$ are not elliptic operators. The coefficient $\mathring{e}^\alpha{}_a$ multiplying the derivative term vanishes at the south pole. We conclude that the choice of the correct mass operator is crucial to establish the spectrum of zero modes after dimensional reduction. This also applies to the usual torsion theories, where often zero modes of the Dirac operator are found, whereas we do not expect that the hermitian operator $\mathring{\mathcal{D}} = \mathring{\Gamma}^\alpha \mathring{D}_\alpha + \frac{1}{2}\tilde{\mathring{L}}^\alpha{}_{\alpha a}$ has any zero modes in most of these cases!

6. Chirality index in generalized gravity

For compact internal spaces the number of chiral fermions is characterized by an index [4] which has been discussed extensively in the context of dimensional reduction by Witten [3]. Let us suppose we have an operator $\bar{D}$ acting on spinors containing first derivatives of the spinor and non-derivative terms and that this operator commutes with the gauge transformations and anticommutes with $\bar{\Gamma}_D$. (D is always even for this section.)

$$[S_z, \bar{D}] = 0, \qquad \{\bar{\Gamma}_D, \bar{D}\} = 0. \tag{122}$$

Examples for such operators are the Dirac operator $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$ or the hermitian operator $\mathring{\mathcal{D}}$. We define [2]

$$\tilde{N}_C(\bar{D}) = (n_C^+ - n_C^- - n_{\bar{C}}^+ + n_{\bar{C}}^-), \tag{123}$$

where n_C^+ counts the number of zero modes of $\bar{D}$ in the Weyl spinor ψ^+ belonging to the complex representation C and similar for n_C^-, $n_{\bar{C}}^+$ etc. ($\bar{C}$ is the complex conjugate representation of C.) If $\bar{D}$ is the mass operator relevant for dimensional reduction (in our case $\mathring{\mathcal{D}}$), the number $\tilde{N}_C$ counts the number of chiral four-dimensional left-handed fermion generations belonging to the representation C, up to a numerical factor to account for eventual Majorana constraints in the $(D+4)$-dimensional theory [2].

If the vielbein $\mathring{e}_\alpha{}^a$ is invertible everywhere, the operators $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$ and $\mathring{\mathcal{D}}$ are elliptic operators. For elliptic operators $\bar{D}$ on compact spaces, the index $\tilde{N}_C(\bar{D})$ remains invariant under continuous changes of the metric, the vielbein and the connections which preserve the isometry group. Although in our examples of sects. 3, 5 the operators $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$ and $\mathring{\mathcal{D}}$ were not elliptic, the index of $\mathring{\Gamma}^\alpha \mathring{D}_\alpha$ was nonzero and the index of $\mathring{\mathcal{D}}$ was zero, independently of the specific r dependence of the functions $f(r)$ and $g(r)$.

For elliptic operators $\overline{D}$ on compact spaces one also has [3]

$$\tilde{N}_C(\overline{D}) = \tilde{N}_C(\overline{D} + tA) \tag{124}$$

for arbitrary t, if A is a non-derivative operator obeying

$$[S_z, A] = 0, \qquad \{\overline{\Gamma}_D, A\} = 0. \tag{125}$$

Examples for operators obeying (125) are $\mathring{U}^{\alpha}{}_{\alpha a}\Gamma^a$ or $\mathring{\tilde{L}}^{\alpha}{}_{\alpha a}\Gamma^a$. Thus $\tilde{N}_C(\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}) = \tilde{N}_C(\mathring{\mathfrak{D}})$ for $\mathring{e}_{\alpha}{}^{a}$ invertible everywhere. In our example, however, the Dirac operator was not elliptic, and this is why we cannot conclude from the nonvanishing index of the Dirac operator $\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha}$ on the existence of chiral massless fermions, which are rather given by the index of $\mathring{\mathfrak{D}}$.

We can use the index to give for compact internal spaces a necessary condition for the existence of chiral fermions in our generalized gravity: The vielbein must not be invertible everywhere. Let us show by steps that an invertible vielbein implies $\tilde{N}_C(\overline{D}) = 0$: we know [3, 4] that for riemannian geometry the chirality index vanishes for the Dirac operator for arbitrary compact spaces with arbitrary isometry group. For ordinary torsion theories $\mathring{e}_{\alpha}{}^{a}\mathring{e}_{\beta a} = \mathring{g}_{\alpha\beta}$, $\mathring{U}_{\alpha\beta a} = 0$, with nonvanishing ground state torsion $\mathring{R}_{\alpha\beta}{}^{\gamma}$ preserving the gauge symmetry ($[\mathring{R}_{\alpha\beta}{}^{\gamma}, S_z] = 0$), one finds from (41) and (71):

$$\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha} = (\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha})_{\mathrm{R}} - \tfrac{1}{2}\mathring{R}_{a}{}^{\alpha}{}_{\alpha}\Gamma^{a} - \tfrac{1}{8}\mathring{R}_{[abc]}\Gamma^{abc}, \tag{126}$$

where $(\mathring{\Gamma}^{\alpha}\mathring{D}_{\alpha})_{\mathrm{R}}$ denotes the Dirac operator of riemannian geometry. By virtue of eq. (125), the Dirac operator has zero index also for nonvanishing torsion. We next release the constraint of vanishing $\mathring{U}_{\alpha\beta a}$, while still keeping $\mathring{e}_{\alpha}{}^{a}\mathring{e}_{\beta a} = \mathring{g}_{\alpha\beta}$. In this case we can again combine (41), (71) and (125) to derive that the Dirac operator has vanishing index. As long as $\mathring{e}_{\alpha}{}^{a}$ is invertible, the Dirac operator still has vanishing index even if we release the constraint $\mathring{e}_{\alpha}{}^{a}\mathring{e}_{\beta a} = \mathring{g}_{\alpha\beta}$. Indeed, we can define a new metric $\mathring{g}_{\alpha\beta}(e) = \mathring{e}_{\alpha}{}^{a}\mathring{e}_{\beta a}$ and the Dirac operator has zero index with respect to the metric $\mathring{g}_{\alpha\beta}(e)$. Now $\mathring{g}_{\alpha\beta}(e)$ can continuously be deformed into the metric $\mathring{g}_{\alpha\beta}$ without breaking the gauge group. Therefore the Dirac operator has vanishing index with respect to $\mathring{g}_{\alpha\beta}$ as well. In conclusion, we find for our generalized gravity that an invertible vielbein on a compact internal space excludes the existence of chiral fermions! It is certainly possible to work out systematically the properties of the index of various operators for the case of vielbeins not invertible everywhere. We do not know if this has been done already. Care has to be taken if the operator is not elliptic. We know no theorem excluding the existence of chiral fermions in this case.

7. Chiral spinors on open internal spaces with finite volume?

So far we have discussed a generalized version of gravity in order to circumvent Witten's no-go theorem for chiral fermions. We considered the internal space to be compact. Compactness of internal space is a basic assumption for the theorem of the vanishing of the chirality index in riemannian geometry. Instead of modifying riemannian geometry, we may investigate noncompact rather than compact spaces.

Indeed, the basic idea for the unification of gauge interactions with gravity in the context of higher-dimensional gravity only assumes that the internal dimensions form a space with a very small characteristic length scale, or equivalently with very strong curvature. We could imagine an open space with finite volume of the order $(M_{\mathrm{P}})^{-D}$. The internal space also should admit a group of isometries which becomes the gauge group after dimensional reduction. Since we observe compact gauge groups, we would like the orbit of the gauge group to be compact. However, the gauge group needs not to act transitively and a compact orbit of the gauge group does not imply that the internal space is compact.

As an example, consider our example of sect. 3, where we now take out the south pole of the manifold. The manifold has then the topology of $\mathbb{R}^D$ and we only need one coordinate patch (eq. (43)). We again assume the ground state to be characterized by a vielbein $\mathring{e}_\alpha{}^a = f(r)\delta_\alpha^a$ and a metric $\mathring{g}_{\alpha\beta} = -g(r)\delta_{\alpha\beta}$ (compare (44)), but we now consider riemannian geometry with $g(r) = f(r)^2$. The function $f(r)$ can be an arbitrary positive definite function. We consider the case where f vanishes faster than r^{-2} for $r \to \infty$. The volume of this space

$$V_D = \int \mathrm{d}^D y\, g_D^{1/2} = \int \mathrm{d}^D y\, f(r)^D \tag{127}$$

turns out to be finite. This gives an example for an open space with finite volume and isometry group SO(D). Assuming the volume to be of the order $(M_{\mathrm{P}})^D$, the characteristic length scale of the internal space is the Planck length and the internal dimensions are therefore not directly observable.

Dimensional reduction in the bosonic sector can be done completely parallel to the case of a compact internal space. The general treatment of the gauge interactions in ref. [11] does not use compactness of the internal space, but rather a finite volume. The gauge coupling is finite provided quantities like

$$\int \mathrm{d}^D y\, g_D^{1/2} \mathring{g}_{\alpha\beta} K_z{}^\alpha K_y{}^\beta \tag{128}$$

are finite. This is the case in our example. Consistency also requires

$$\int \mathrm{d}^D y\, g_D^{1/2} R_{\mathrm{I}} \tag{129}$$

to be finite, with R_{I} the internal curvature scalar. For more complicated effective actions including terms like $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ [13], the integral over these terms should be finite. We then expect finite gauge coupling and Newton's constant, normalizable kinetic terms (up to a wrong sign if the "ground state" is unstable) and a well-defined scalar potential. We will not address in this paper the question if such a noncompact internal space could be an acceptable ground state, which means that it is at least a classically stable solution of the field equations derived from some appropriate effective action for d-dimensional gravity. We are first interested in the question if such an open internal space with finite volume could admit chiral fermions.

Let us pursue our example with SO(D) symmetry: the relevant geometrical quantities are

$$\mathring{e}_\alpha{}^a = f(r)\mathring{\delta}^a_\alpha\,, \qquad \mathring{g}_{\alpha\beta} = -f^2(r)\delta_{\alpha\beta}\,,$$

$$\mathring{\omega}_{abc} = \frac{y^\gamma}{r}\frac{f'}{f^2}\left(\delta_{\gamma b}\delta_{ac} - \delta_{\gamma c}\delta_{ab}\right). \tag{130}$$

One finds a vanishing totally antisymmetric part of the spin connection

$$\tilde{\omega}_{abc} = \omega_{[abc]} = 0\,, \tag{131}$$

and the Dirac operator is given by

$$\mathring{\Gamma}^\alpha\mathring{D}_\alpha = \mathring{\mathfrak{D}} = \Gamma^a f^{-1}\left(\delta^\alpha_a\partial_\alpha + \tfrac{1}{2}(D-1)\frac{y^\alpha}{r}\frac{f'}{f}\delta_{\alpha a}\right). \tag{132}$$

We note that this is the same operator as in sect. 5, eq. (10) for $g = f^2$. This is an example how two different spin connections with the same totally antisymmetric part $\tilde{\omega}_{abc}$ lead to the same mass operator.

We consider a harmonic expansion of a Dirac spinor

$$\psi(y) = a_{nHMj}\psi_{nHMj}(y)\,,$$

$$\mathring{\mathfrak{D}}\psi_{nHMj} = M\psi_{nHMj}\,, \tag{133}$$

where we restrict our basis functions to functions with vanishing divergence

$$\int \mathrm{d}^D y\,\partial_\alpha\left(g_D^{1/2}\psi_{nHMj}{}^\dagger\Gamma^\alpha\psi_{nHMj}\right) = 0\,. \tag{134}$$

The functions ψ_{nHMj} are obtained as solutions of the equation

$$\mathring{\mathfrak{D}}^2\psi_{nHMj} = M^2\psi_{nHMj}\,, \tag{135}$$

with positive definite or vanishing M^2. We expect the spectrum of $\mathring{\mathfrak{D}}^2$ to be discrete.

We first specialize to the case $D=2$ (with notation as in sect. [5]). Zero modes of $\mathring{\mathfrak{D}}$ obey again eq. (116) with solution (117). However, the question of normalizability is now different, since $g(r)$ vanishes faster than r^{-4} for $r \to \infty$. The relevant integral (118) is given by

$$I_n \sim \int_0^{\infty} \mathrm{d}r\, r^{2n+1} f(r). \tag{136}$$

A finite value of f at $r=0$ implies again that normalizable zero modes exist only for $n \geqslant 0$. Assuming

$$\lim_{r \to \infty} f(r) \to a r^{-N} \tag{137}$$

we find a normalizable zero mode in χ^{+} for $n=0$ provided $2<N \leqslant 4$. This corresponds to a chiral spinor in ψ^{+} with charge $+\frac{1}{2}$. The charge conjugate of this spinor is also massless. It is in ψ^{-} and has charge $-\frac{1}{2}$. For higher values of N we find even more chiral spinors. For $4<N \leqslant 6$ there are normalizable spinors for $n=0$ and $n=1$, and in general normalizable zero modes for $0 \leqslant n \leqslant M$ exist for $2(M+1)<N \leqslant 2(M+2)$. We have found an example of a nonvanishing chirality index of the mass operator $\mathring{\mathfrak{D}}$.

The $n=0$ zero mode is easily generalized to more than two internal dimensions. We make an ansatz for a zero-mass solution:

$$\psi_0 = \varphi(r) \chi, \tag{138}$$

with χ a constant spinor and $\varphi(r)$ a real function of r. Since $K_{AB}{}^{\alpha} \partial_{\alpha} \varphi(r)=0$, the spinor ψ_0 has the same transformation properties as the constant spinor χ in sect. 3:

$$\delta \psi_0 = \tfrac{1}{2} \theta^{AB} \Sigma_{AB} \psi_0 . \tag{139}$$

We remind ourselves that for $D = 2 \bmod 4$ ψ_0^{+} and ψ_0^{-} belong to complex spinor representations conjugate to each other. Solutions of the equation

$$\mathring{\mathfrak{D}} \psi_0 = 0 \tag{140}$$

obey

$$\frac{\varphi'}{\varphi} + \tfrac{1}{2}(D-1) \frac{f'}{f} = 0 \tag{141}$$

and are given by

$$\varphi(r) = \varphi_0 f(r)^{-(D-1)/2} \tag{142}$$

The zero mode $\varphi(r)$ is normalizable if the integral

$$I_n \sim \int_0^\infty \mathrm{d}r\, r^{D-1} f(r) \tag{143}$$

is finite. Assuming again that $f(r)$ vanishes like r^{-N} for large r (eq. 137), we find a normalizable zero mode provided $N > D$. If there are no other normalizable zero modes, we again have found an example of chiral fermions. (By analogy with the two-dimensional example, we do not expect supplementary zero modes if f does not vanish much faster than $r^{-(D+1)}$.) Starting from $d = 14$ riemannian gravity, a small noncompact 10-dimensional internal space would lead to two identical generations of left-handed fermions belonging to the 16-dimensional spinor representation of SO(10)! The existence of two identical fermion generations is related to the fact that there exists no Majorana spinor in 14-dimensional gravity [10]. To make the picture more realistic, one would be tempted to start with a Majorana-Weyl spinor in $d = 18$ and use the supplementary four internal dimensions to provide a generation group for possibly four generations of chiral fermions.

We do not know yet if the approach of a noncompact internal space will lead to a consistent theory. Many questions remain open. For example, the curvature tensor in our example is given by

$$\begin{aligned}\mathring{R}_{\alpha\beta ab} = {} & \left(\frac{2f'}{rf} + \frac{f'^2}{f^2}\right)(\delta_{\alpha a}\delta_{\beta b} - \delta_{\beta a}\delta_{\alpha b}) \\ & + \frac{y^\gamma y^\delta}{r^2}\left(\frac{f''}{f} - \frac{f'}{rf} - 2\frac{f'^2}{f^2}\right)(\delta_{\gamma a}\delta_{\beta b}\delta_{\delta\alpha} - \delta_{\gamma b}\delta_{\beta a}\delta_{\delta\alpha} - \delta_{\gamma a}\delta_{\alpha b}\delta_{\delta\beta} + \delta_{\gamma b}\delta_{\alpha a}\delta_{\delta\beta})\end{aligned} \tag{144}$$

with curvature scalar

$$\begin{aligned}\mathring{R} &= \mathring{e}^{a\alpha}\mathring{e}^{b\beta}\mathring{R}_{\alpha\beta ab} \\ &= (D-1)\left\{D\left(\frac{2f'}{rf^3} + \frac{f'^2}{f^4}\right) + 2\left(\frac{f''}{f^3} - \frac{f'}{rf^3} - 2\frac{f'^2}{f^4}\right)\right\}.\end{aligned} \tag{145}$$

For $f(r)$ vanishing for large r like ar^{-N}, $N > 2$, the curvature scalar $\mathring{R}$ diverges for $r \to \infty$ like

$$\lim_{r\to\infty} \mathring{R} \to \frac{(D-1)N}{a^2} r^{2(N-2)}\{(D-2)(N-2)r^2 - 2D\}. \tag{146}$$

However, the integral

$$\int \mathrm{d}^D y\, g_D^{1/2}(\mathring{R})^M \sim \int \mathrm{d}r\, r^{(2M-D)(N-1)-1} \tag{147}$$

is finite for all $M < \frac{1}{2}D$. We do not know if an acceptable ground state admitting chiral fermions can be found. (Questions of classical stability of a possible ground state have been discussed extensively in ref. [6].) Also a more systematic study of the harmonic expansion is needed.

We should not be too afraid of the divergences in geometrical quantities or in the zero mode solutions for spinors as r goes to infinity. For large r the characteristic length scale becomes extremely tiny since the curvature scalar diverges for $r \to \infty$. After all, our knowledge about the physics at very short distances is very poor. Problems occurring at extremely short characteristic length scales may be related to the need of a more complete physical theory at these distances. We have neglected so far quantum fluctuations in our completely classical treatment of the theory.

The conclusion of this paper can be made rather short: there are still interesting possibilities left open, how pure higher-dimensional gravity could lead to chiral spinors in four dimensions. They certainly merit further investigation.

References

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys. K1 (1921) 966;
O. Klein, Z. Phys. 37 (1926) 895;
A. Einstein and P. Bergmann, Ann. Math. 39 (1938) 683;
B. De Witt, in Relativity, groups and topology (Gordon and Breach, New York, 1964);
J. Rayski, Acta Phys. Pol. 27 (1965) 89;
R. Kerner, Ann. Inst. Henri Poincaré 9 (1968) 143;
A. Trautmann, Reports Math. Phys. 1 (1970) 29;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1975) 1711;
L.N. Chang, K.I. Macrae and F. Mansouri, Phys. Rev. D13 (1976) 235;
J. Scherk and J.H. Schwarz, Phys. Lett. 57B (1975) 463; Nucl. Phys. B153 (1979) 61;
E. Cremmer and J. Scherk, Nucl. Phys. B103 (1976) 393; B108 (1976) 409;
J.F. Luciani, Nucl. Phys. B135 (1978) 111;
E. Witten, Nucl. Phys. B186 (1981) 412;
A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316

[2] C. Wetterich, Nucl. Phys. B223 (1983) 109

[3] E. Witten, Princeton preprint (1983)

[4] M.F. Atiyah and F. Hirzebruch, in Essays on topology and related topics, ed. A. Haefliger and R. Narasimhan (Springer, Berlin, 1970) p. 18

[5] N.S. Manton, Nucl. Phys. B193 (1981) 391;
G. Chapline and N. Manton, Nucl. Phys. B184 (1981) 391;
G. Chapline and R. Slansky, Nucl. Phys. B209 (1982) 461

[6] S. Randjbar-Daemi, A. Salam and J. Strathdee, Nucl. Phys. B214 (1983) 491; Phys. Lett. 124B (1983) 349; Trieste preprint (1983)

[7] S. Weinberg, Phys. Lett. 138B (1984) 47

[8] C. Wetterich, CERN preprint TH3528, revised version (1983), to appear in Phys. Lett. B

[9] C. Wetterich, Nucl. Phys. B211 (1983) 177

[10] C. Wetterich, Nucl. Phys. B222 (1983) 20

[11] C. Wetterich, Phys. Lett. 110B (1982) 379

[12] C. Destri, C. Orzalesi and P. Rossi, Ann. Phys. 147 (1983) 321;
Y.S. Wu and A. Zee, Univ. of Washington preprint 400 48-25P3 (1983)

[13] C. Wetterich, Phys. Lett. 113B (1982) 377

SPONTANEOUS COMPACTIFICATION IN SIX-DIMENSIONAL EINSTEIN–MAXWELL THEORY

S. RANDJBAR-DAEMI[1]

International Centre for Theoretical Physics, Trieste, Italy

ABDUS SALAM

International Centre for Theoretical Physics, Trieste, Italy,
and
Imperial College, London, England

J. STRATHDEE

International Centre for Theoretical Physics, Trieste, Italy

Received 8 November 1982

A discrete set of solutions to the classical Einstein–Maxwell equations in six-dimensional space-time is considered. These solutions have the form of a product of four-dimensional constant curvature space-time with a 2-sphere. The Maxwell field has support on the 2-sphere where it represents a monopole of magnetic charge, $n = \pm 1, \pm 2, \ldots$. The spectrum of massless and massive states is obtained for the special case of flat 4-space, and the solution is shown to be classically stable. The limiting case where the radius of the 2-sphere becomes small is considered and a dimensionally reduced effective lagrangian for the long range modes is derived. This turns out to be an SU(2)×U(1) gauge theory with chiral couplings.

1. Introduction

To illustrate the phenomenon of spontaneous compactification in theories of the Kaluza–Klein type, a six-dimensional model was constructed by Horvath, Palla, Cremmer and Scherk [1, 2]. In this model the gravitational field is coupled to an abelian gauge field*. Three independent parameters are involved, the gravitational and gauge coupling constants and a cosmological constant. Among the possible solutions it was shown that there occurs a candidate ground state in which the geometry factorizes into the product of a four-dimensional space–time of constant curvature with a two-dimensional sphere. In this solution the Maxwell field is non-vanishing only on the 2-sphere where it assumes a magnetic monopole configuration. This non-vanishing ground state value of the Maxwell field contributes to the stress-energy tensor on the right-hand side of the six-dimensional Einstein

[1] Address after 1st December 1982: Institute for Theoretical Physics, University of Vienna, Boltzmanngasse 5, Vienna, Austria.

* A class of non-abelian generalizations has recently been obtained [3].

equations and, in effect, generates the curvature of the 2-sphere. In view of the non-zero "magnetic" charge of the proposed vacuum configuration, it was conjectured that the solution would be a stable one. One of our aims in this article is to test the solution for stability against small perturbations.

Our main purpose, however, is to examine in some detail the Kaluza–Klein mechanism which operates in this model. In particular, we shall compute the excitation spectrum for the special case where the four-dimensional spacetime is flat. It will be shown that the spectrum comprises several series of massive states with spins 0, 1 and 2, and ten massless states: the graviton (helicities ±2) and gauge vectors (helicities ±1) corresponding to the local symmetry $SU(2)\times U(1)$. Perhaps surprisingly, there are no massless scalars.

A secondary aim is to illustrate the use of harmonic expansions [4] on the 2-sphere in analyzing the spectrum. This is not entirely trivial and, indeed, we shall find, contrary to the usual assumption, that the SU(2) Yang–Mills 4-vector does not lie entirely in the six-dimensional metric, but is partly in the Maxwell 6-vector as well. In addition, we shall find that its coupling to massless fermions is chiral*.

An important feature of this model is the presence of two independent scales. In the vacuum geometry the curvature of the 2-sphere can be varied independently of the four-dimensional spacetime curvature. This enables us to consider the special case where spacetime is flat** and it further enables us to extract in a meaningful way the effective four-dimensional theory of the zero modes by allowing the radius of the 2-sphere to shrink to zero***

If the four-dimensional curvature is not adjusted to zero, the candidate ground state would have either de Sitter ($\Lambda > 0$) or anti-de Sitter ($\Lambda < 0$) geometry. It has been shown by Sezgin [6], using the approach of Abbot and Deser [7] that the anti-de Sitter solution is stable in that a conserved, non-negative, energy functional can be defined. Here we consider only the flat case because the interpretation of particle spectra is more straightforward.

This paper is arranged as follows. Sect. 2 sets out the model and the possible ground state solutions are reviewed. Sect. 3 is devoted to the fluctuation spectrum. Fields are expanded around their background values and the linearized equations of motion are obtained. These are solved in the presence of external sources in a suitable gauge. By this means the fluctuation Green functions are in effect obtained. The singularities of these Green functions define the spectrum. It is verified that no tachyons or ghosts are contained. For this investigation it is necessary to distinguish physical states from gauge modes and we do this by using appropriately conserved sources. Dimensional reduction is considered in sect. 4 where an effective lagrangian for the zero modes is obtained by shrinking the 2-sphere. Fermions are introduced in sect. 5 where it is shown that zero modes occur whose SU(2) quantum numbers are governed by the Chern class of the monopole background.

---

* A general discussion of the chirality question can be found in ref. [3a].

** This is to be contrasted with the interesting solution of Freund and Rubin [5] to the 11-dimensional supergravity system where the four-dimensional anti-de Sitter space has curvature of comparable magnitude to that of the "internal" 7-space.

***It should be observed, however, that the shrinking of the 2-sphere leads to a regime of strong couplings for the gauge fields. The classical approximation which we employ cannot be reliable in this limit.

2. Ground state background

The six-dimensional Einstein–Maxwell theory with cosmological constant is characterized by the action

$$S=-\int \mathrm{d}^6z\,\sqrt{-g}\left[\frac{1}{\kappa^2}R+\tfrac{1}{4}F_{MN}F^{MN}+\lambda\right], \tag{2.1}$$

where R denotes the curvature scalar*, and

$$F_{MN}=\partial_M A_N-\partial_N A_M\,. \tag{2.2}$$

The field equations are

$$R_{MN}-\tfrac{1}{2}g_{MN}R=-\tfrac{1}{2}\kappa^2(T_{MN}-\lambda g_{MN})\,, \tag{2.3}$$

$$\frac{1}{\sqrt{-g}}\,\partial_M(\sqrt{-g}F^{MN})=0\,, \tag{2.4}$$

where the stress energy tensor takes the usual Maxwell form,

$$T_{MN}=F_{ML}F_N{}^L-\tfrac{1}{4}g_{MN}F^2\,. \tag{2.5}$$

The most symmetric solution to the eqs. (2.3) and (2.4) would of course be six-dimensional Minkowski space with vanishing Maxwell field. However, such a solution could hardly be relevant to the description of a four-dimensional world. We shall therefore restrict our attention to the most symmetric solutions with the structure $M^4\times M^2$, viz. the product of four- and two-dimensional spaces of constant curvature. Thus, we take

$$g_{MN}\,\mathrm{d}z^M\,\mathrm{d}z^N=g_{mn}(x)\,\mathrm{d}x^m\,\mathrm{d}x^n+g_{\mu\nu}(y)\,\mathrm{d}y^\mu\,\mathrm{d}y^\nu\,, \tag{2.6}$$

$$A_M\,\mathrm{d}z^M=A_\mu(y)\,\mathrm{d}y^\mu\,, \tag{2.7}$$

where $g_{mn}(x)$ corresponds to de Sitter or anti-de Sitter space and $g_{\mu\nu}(y)$ to a 2-sphere. The 1-form, $A_\mu\,\mathrm{d}y^\mu$ is required to be invariant (up to a gauge transformation) under rotations of the 2-sphere. We believe that physical interpretations are less vexed in the anti-de Sitter case (having a global time-like Killing vector and positive energy unitary representations) than they are for the de Sitter world and we shall see that it is possible to exclude the de Sitter solution by appropriate restrictions on the parameters κ, λ and e. Moreover, it can be shown that the anti-de Sitter solution is classically stable [6].

---

* Upper case indices take the values 0, 1, 2, 3, 5, 6. In the following, lower case latin indices will take the values 0, 1, 2, 3, while lower case greek indices are restricted to 5, 6. That is, we write $z^M=(x^m, y^\mu)$. Early alphabet letters will be used as frame labels, while the mid-alphabet is used for world indices. We follow the conventions of Deser and van Nieuwenhuizen [8]: The signature is $-+\cdots+$. The Riemann tensor is defined such that $R^K_{\ LMN}=\partial_M\Gamma^K_{\ LN}-\partial_N\Gamma^K_{\ LM}-\cdots$, and the Ricci tensor is $R_{LM}=R^K_{\ LMK}$.

It is convenient to use spherical polar coordinates on the 2-sphere,

$$g_{\mu\nu}\,\mathrm{d}y^{\mu}\,\mathrm{d}y^{\nu}=a^{2}(\mathrm{d}\theta^{2}+\sin^{2}\theta\,\mathrm{d}\varphi^{2})\,, \tag{2.8}$$

where a, the radius is to be determined. It is *not* an integration constant.

The rotation invariant Maxwell field is given by

$$A_{\mu}(y)\,\mathrm{d}y^{\mu}=\frac{n}{2e}(\cos\theta\mp 1)\,\mathrm{d}\varphi\,, \tag{2.9}$$

where n is a positive or negative integer and e is the U(1) gauge coupling constant. This coupling constant does not appear in the lagrangian (2.1), where all fields are neutral, but one should keep in mind the possibility of charged matter fields to which A would couple. An example of this will be dealt with in sect. 5. In effect we are asking that the theory should be invariant under the gauge transformations

$$A_{M}\rightarrow A_{M}+\frac{1}{e}\partial_{M}\Lambda\,,$$

where $\exp(i\Lambda)$ should be single-valued. In (2.9) the monopole configuration is necessarily specified on two patches ($0\leqslant\theta<\pi$ and $0<\theta\leqslant\pi$, $0\leqslant\varphi<2\pi$, respectively). Where these neighbourhoods overlap, say at $\theta=\pi/2$, $0\leqslant\varphi\leqslant 2\pi$, the two representations of A must be related by a single-valued gauge transformation, $\exp(in\varphi)$. Hence the requirement that n be a positive or negative integer. The field strength corresponding to (2.9) is given by the 2-form,

$$F=\mathrm{d}A=-\frac{n}{2ea^{2}}a\,\mathrm{d}\theta\wedge a\sin\theta\,\mathrm{d}\varphi\,, \tag{2.10}$$

where a is the radius of the 2-sphere. It is clearly rotation invariant.

Corresponding to the field strength (2.10) the components of the stress energy tensor (2.5), in an orthonormal basis, are given by

$$\begin{aligned}
&T_{ab}=-\frac{n^{2}}{8e^{2}a^{4}}\eta_{ab}\,, \qquad a,b=0,1,2,3\,,\\
&T_{a\beta}=0\,,\\
&T_{\alpha\beta}=\frac{n^{2}}{8e^{2}a^{4}}\delta_{\alpha\beta}\,, \qquad \alpha,\beta=5,6\,.
\end{aligned} \tag{2.11}$$

The non-vanishing parts of the Riemann tensor, in the same basis are,

$$R_{abcd} = \frac{\kappa^2 \Lambda}{6}(\eta_{ac}\eta_{bd} - \eta_{ad}\eta_{bc}) ,$$
$$R_{\alpha\beta\gamma\delta} = \frac{1}{a^2}(\delta_{\alpha\gamma}\delta_{\beta\delta} - \delta_{\alpha\delta}\delta_{\beta\gamma}) , \tag{2.12}$$

where the four-dimensional cosmological constant, Λ, is to be determined along with the radius a, by substituting (2.12) into the Einstein equations (2.3), or the equivalent form

$$R_{AB} = -\frac{\kappa^2}{2}(T_{AB} - \tfrac{1}{4}\eta_{AB}T + \tfrac{1}{2}\eta_{AB}\lambda) .$$

One finds two algebraic equations for Λ and a^2,

$$\Lambda = -\frac{n^2}{16e^2a^4} + \frac{\lambda}{2}, \qquad \frac{2}{\kappa^2 a^2} = \frac{3n^2}{16e^2a^4} + \frac{\lambda}{2} . \tag{2.13}$$

They are solved by

$$\frac{1}{a^2} = \frac{16e^2}{3n^2\kappa^2}\left[1 \pm \sqrt{1 - \frac{3n^2}{32}\frac{\kappa^4\lambda}{e^2}}\right], \tag{2.14}$$

$$\Lambda = \tfrac{2}{3}\lambda - \frac{32}{9n^2}\frac{e^2}{\kappa^4}\left[1 \pm \sqrt{1 - \frac{3n^2}{32}\frac{\kappa^4\lambda}{e^2}}\right] . \tag{2.15}$$

If $\lambda < 0$ then, for each value of the monopole charge, n, there is one positive solution for a^2. This solution corresponds to $\Lambda < 0$ (anti-de Sitter world). On the other hand, if $\lambda > 0$, then there are two positive solutions for a^2, provided n^2 is not too large,

$$n^2 \leqslant \frac{32}{3}\frac{e^2}{\kappa^2\lambda} .$$

At the upper end of this range, however, one finds $\Lambda > 0$ (de Sitter world) which should probably be excluded. To have $\Lambda \leqslant 0$ we require

$$n^2 \leqslant 8\frac{e^2}{\kappa^4\lambda} . \tag{2.16}$$

Of particular interest for our purposes is the case of flat 4-space, $\Lambda = 0$. This occurs when the parameters of the six-dimensional theory are adjusted such that

$$e^2 = \tfrac{1}{8}n^2\kappa^4\lambda , \tag{2.17}$$

and the radius then takes the value

$$a^2 = \tfrac{1}{8}n^2\frac{\kappa^2}{e^2} . \tag{2.18}$$

We shall examine the fluctuations on this background in sect. 3.

We conclude by remarking that the four-dimensional theory based on this vacuum will be invariant under the transformations of Poincaré × $(\mathrm{SU}(2)\times \mathrm{U}(1))_{\mathrm{local}}$. The local SU(2) corresponds to x-dependent rotations of the 2-sphere with appropriate U(1) transformations in the tangent space as explained in ref. [4]. The local U(1) corresponds to the four-dimensional part of the Maxwell gauge symmetry.

3. Fluctuation spectrum ($\Lambda = 0$)

The classical test for stability would be to compute the total energy carried by fluctuations on the background solution and show that it is positive. Such an analysis has been carried out for $\Lambda \leq 0$ and these solutions are indeed stable, at least for weak fluctuations [6]. Our purpose here is somewhat different in that we would like to see what kinds of states are produced in this theory. In particular, we shall derive the spectrum of masses for the scalar, vector and tensor excitations on the $\Lambda = 0$ background. In the course of this we shall verify that there are no tachyons or negative metric ghosts: the background is therefore at least perturbatively stable.

In order to obtain the spectrum one must substitute into the action functional (2.1) the expressions

$$g_{MN} = \bar{g}_{MN} + \kappa h_{MN}\,, \qquad A_M = \bar{A}_M + V_M\,, \tag{3.1}$$

where $\bar{g}_{MN}$, $\bar{A}_M$ denote one of the ground state solutions obtained above, and h_{MN}, V_M represent the fluctuations. Terms linear in h, V will vanish by virtue of the field equations satisfied by $\bar{g}$, $\bar{h}$. Terms bilinear in h, V will govern the propagation of the weak disturbances. These we retain. Higher order terms, representing interactions will be discarded. In an orthonormal basis the bilinear terms reduce to

$$\begin{aligned} S_2 = \int \mathrm{d}^6 z\, \sqrt{-\bar{g}}\Big[& -\tfrac{1}{4} h_{AB;C} h_{AB;C} + \tfrac{1}{8} h_{AA;C} h_{BB;C} \\ & + \frac{1}{4}\left(\bar{R} + \frac{\kappa^2}{4}\bar{F}^2 + \kappa^2 \lambda\right)(h_{AB}h_{AB} - \tfrac{1}{2} h_{AA}h_{BB}) \\ & - \tfrac{1}{2}\bar{R}_{BC}(h_{AB}h_{AC} - h_{AA}h_{BC}) \\ & - \frac{\kappa^2}{2}\bar{F}_{BD}\bar{F}_{CD}(h_{AB}h_{AC} - \tfrac{1}{2}h_{AA}h_{BC}) \\ & + \frac{1}{2}\left(R_{ABCD} - \frac{\kappa^2}{2}\bar{F}_{AC}\bar{F}_{BD}\right)h_{AB}h_{CD} \\ & - \tfrac{1}{2}V_{A;B}V_{A;B} + \tfrac{1}{2}\bar{R}_{AB}V_A V_B \\ & + \kappa(h_{AB}\bar{F}_{BC} - \tfrac{1}{4}h_{BB}\bar{F}_{AC})(V_{C;A} - V_{A;C}) \\ & + \tfrac{1}{2}T_{AB}h_{AB} + J_A V_A\Big]\,, \end{aligned} \tag{3.2}$$

where external sources T_{AB} and J_A have been included. The semicolon notation indicates covariant derivatives using the background connection. In general, the bilinear part of the action would reflect the general co-ordinate invariance and U(1) gauge invariance of the original action as invariance under the transformations

$$\delta h_{AB} = \frac{1}{\kappa}(\xi_{A;B} + \xi_{B;A}), \qquad \delta V_A = \omega_{;A} - \bar{F}_{AB}\xi_B . \tag{3.3}$$

We have taken account of this by imposing the gauge conditions

$$h_{AB;B} = \tfrac{1}{2}h_{BB;A}, \qquad V_{A;A} = 0 . \tag{3.4}$$

It should now be observed that the linear equations of motion for h, V obtained from (3.2) are compatible only if the sources are suitably constrained, viz.

$$T_{AB;B} = \kappa \bar{F}_{AB} J_B, \qquad J_{A;A} = 0 . \tag{3.5}$$

In a complete theory these conservation laws would follow from the equations of motion of those fields which contribute to the sources. Here we must view them as a necessary characterization of "physical" sources (i.e. on-shell matrix elements of current operators).

The next step is to solve the linear equations for h and V in terms of T and J. The solutions are then substituted into the functional (3.2). The result reduces to

$$I(T,J) = \tfrac{1}{2}\int \mathrm{d}^6 z\,\sqrt{-\bar{g}}\,[\tfrac{1}{2}T_{AB}h_{AB} + J_A V_A], \tag{3.6}$$

where h and V are functional of J and T. This functional defines the components of the propagator,

$$S_2 = \tfrac{1}{2}\int \mathrm{d}^6 z \sqrt{-\bar{g}}\,\mathrm{d}^6 z'\,\sqrt{-\bar{g}'}$$
$$\times (T_{AB}(z) J_A(z)) \begin{pmatrix} \tfrac{1}{4}\langle h_{AB}(z) h_{CD}(z')\rangle & \tfrac{1}{2}\langle h_{AB}(z) V_C(z')\rangle \\ \tfrac{1}{2}\langle V_A(z) h_{CD}(z') & \langle V_A(z) V_C(z')\rangle \end{pmatrix} \begin{pmatrix} T_{CD}(z') \\ J_C(z') \end{pmatrix} . \tag{3.7}$$

In practice, the computation would be rather awkward because of the curved background associated with the 2-sphere. To cope with this we have found it useful to expand all fields and sources in spherical harmonics on the 2-sphere. The invariance of the sphere under SO(3) implies that the propagators will be diagonal in an isospin-like quantum number $l = 0, 1, 2, \ldots$ In particular, poles of the four-dimensional Fourier transform will be labelled by l. We shall obtain in this way several sequences of masses, the excitation spectrum.

The expansion method is discussed in detail in ref. [4]. Here, in order to fix the notation, we summarize the main features. To begin, viewing the 2-sphere as a quotient space, SU(2)/U(1) introduce the SU(2) boosts, $L_{\theta\varphi}$, to parametrize it,

$$L_{\theta\varphi} = \mathrm{e}^{-\varphi Q_3}\,\mathrm{e}^{-\theta Q_2}\,\mathrm{e}^{\varphi Q_3} , \tag{3.8}$$

(excluding* $\theta = \pi$) where $[Q_1, Q_2] = Q_3$, etc. The action of an SU(2) element, g, on this space is defined by

$$gL_{\theta\varphi} = L_{\theta'\varphi'}h\,, \tag{3.9}$$

where $h = \exp(\zeta Q_3)$ depends on θ, φ and g. The 1-form, $L^{-1}\,\mathrm{d}L$, belongs to the algebra of SU(2) and its components comprise a covariant basis, $E^{\pm}$, and the U(1) connection 1-form, e^3,

$$L^{-1}\,\mathrm{d}L = \frac{1}{a}E^{+}\frac{Q_1 - iQ_2}{\sqrt{2}} + \frac{1}{a}E^{-}\frac{Q_1 + iQ_2}{\sqrt{2}} + e^3 Q_3\,, \tag{3.10}$$

where

$$\begin{aligned} E^{\pm} &= \pm\frac{a}{i\sqrt{2}}\,\mathrm{e}^{\mp i\varphi}\,(\mathrm{d}\theta \mp i\sin\theta\,\mathrm{d}\varphi)\,, \\ e^3 &= -\mathrm{d}\varphi(\cos\theta - 1). \end{aligned} \tag{3.11}$$

The frames $E^{\pm}$ are related to those of sect. 2 by

$$E^{\pm} = \frac{1}{\sqrt{2}}(E^5 \pm iE^6)\,. \tag{3.12}$$

From (3.9) and (3.10) one finds that, under the action of SU(2), the frames and connection transform according to

$$\begin{aligned} E^{\pm}(y) &\to E^{\pm}(y') = E^{\pm}(y)\,\mathrm{e}^{\pm i\zeta}\,, \\ e^3(y) &\to e^3(y') = e^3(y') - \mathrm{d}\zeta\,. \end{aligned} \tag{3.13}$$

Defining the covariant differential of $L^{-1}_{\theta\varphi}$,

$$\begin{aligned} \nabla L^{-1} &= \mathrm{d}L^{\;\;1} + e^3 Q_3 L^{\;-1} \\ &= E^{+}\nabla_{+}L^{-1} + E^{-}\nabla_{-}L^{-1}\,, \end{aligned} \tag{3.14}$$

---

* Strictly, one should employ two patches. The parametrization (3.8) is suitable for $0 \leqslant \theta < \pi$, $0 \leqslant \varphi < 2\pi$. For the other patch, $0 < \theta \leqslant \pi$, $0 \leqslant \varphi < 2\pi$ one should write $L_{\theta\varphi} = \mathrm{e}^{-\varphi Q_3}\,\mathrm{e}^{-\theta Q_2}\,\mathrm{e}^{-\varphi Q_3}$. The action of $L_{\theta\varphi}$ can be seen by embedding S^2 in R^3 where spherical polar co-ordinates give the parametrization

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{pmatrix} = D(L_{\theta\varphi})\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

where

$$D(L_{\theta\varphi}) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

one finds, by simply rearranging (3.10), the useful formula

$$\nabla_{\pm} L^{-1} = -\frac{1}{a}\frac{Q_1 \mp iQ_2}{\sqrt{2}} L^{-1} . \qquad (3.15)$$

The behaviour (3.13), in conjunction with the identification (3.12) means that rotations of the 2-sphere must be associated with tangent space rotations in the 5-6 plane in order to preserve the form of the background (2.9). Observe that the background Maxwell field is proportional to $e^3(y)$. Therefore, fields to be expanded must first be decomposed into irreducible representations of these SO(2) rotations, labelled by the "iso-helicity", λ. Let $\phi_\lambda(x, \theta, \varphi)$ be a typical one. It can then be shown that the harmonic expansion of ϕ_λ takes the form

$$\phi_\lambda(x, \theta, \varphi) = \sum_{l \geqslant |\lambda|} \sqrt{2l+1} \sum_m D^l_{\lambda m}(L^{-1}_{\theta\varphi}) \phi^l_{\lambda m}(x) , \qquad (3.16)$$

where $D^l_{\lambda m}$ belongs to the $2l+1$-dimensional unitary irreducible representation of SU(2).

By virtue of the formula (3.15) the computation of covariant derivatives reduces to algebraic manipulation*,

$$\begin{aligned}\nabla_{\pm} D^l_{\lambda m}(L^{-1}) &= -\frac{1}{a\sqrt{2}} D^l_{\lambda m}(Q_{1\mp i2} L^{-1}) \\ &= +\frac{i}{a}\sqrt{\tfrac{1}{2}(l \mp \lambda)(l \pm \lambda + 1)}\, D^l_{\lambda \pm 1, m}(L^{-1}) . \qquad (3.17)\end{aligned}$$

This feature leads to considerable simplications in the equations of motion.

Among the fields h_{AB} and V_A we have the following SO(2) irreducible pieces:

$$\begin{aligned} h_{++} &= \tfrac{1}{2}(h_{55} - h_{66} - 2ih_{56}) , && \lambda = 2 , \\ h_{a+} &= \sqrt{\tfrac{1}{2}}(h_{a5} - ih_{a6}) , && \lambda = 1 , \\ h_{+-} &= \tfrac{1}{2}(h_{55} + h_{66}) , && \lambda = 0 , \\ V_+ &= \sqrt{\tfrac{1}{2}}(V_5 - iV_6) , && \lambda = 1 , \qquad (3.18) \end{aligned}$$

plus h_{ab}, V_a with $\lambda = 0$ and the complex conjugates $h_{--} = h^*_{++}$, etc. Each of these has an expansion of the form (3.16). The sources, T_{AB} and J_A are likewise expanded.

---

* We are using anti-hermitian generators, Q, with phase conventions [9] such that

$$\langle\lambda| Q_{1-i2} = \frac{1}{i}\sqrt{(l-\lambda)(l+\lambda+1)}\,\langle\lambda+1| ,$$

$$\langle\lambda| \mathrm{e}^{\zeta Q_3} = \mathrm{e}^{-i\lambda\zeta}\langle\lambda| .$$

Under the tangent space U(1) associated with the global action of an element, g, of SU(2) on S^2,

$$\phi'_\lambda(x, \theta', \varphi') = \mathrm{e}^{-i\lambda\zeta}\phi_\lambda(x, \theta, \varphi) .$$

This implies invariance of $\mathrm{E}^+\phi_+$, etc., as it should.

The equations of motion for h and V derived from (3.2) are

$$\begin{aligned}
&\nabla^2(h_{AB}-\tfrac{1}{2}\eta_{AB}h_{CC})+(\bar{R}_{AC}h_{CB}+\bar{R}_{BC}h_{CA})\\
&\quad+\kappa\bar{F}_{BC}(\nabla_A V_C-\nabla_C V_A)+\kappa\bar{F}_{AC}(\nabla_B V_C-\nabla_C V_B)-\kappa\eta_{AB}\bar{F}_{CD}\nabla_C V_D=-T_{AB}\,,\\
&\qquad\nabla^2 V_A+R_{AB}V_B+\kappa\nabla_C(h_{AB}\bar{F}_{BC}+\bar{F}_{AB}h_{BC}+\tfrac{1}{2}h_{BB}\bar{F}_{CA})=-J_A\,.
\end{aligned}\tag{3.19}$$

Considerable simplifications have been brought about in these equations by use of the background solution. The only non-vanishing components of $\bar{R}$ and $\bar{F}$ are, in the basis (3.11),

$$\bar{R}_{+-}=\bar{R}_{-+}=-\frac{1}{a^2}\,,\qquad \bar{F}_{+-}=-\bar{F}_{-+}=\frac{i}{a}\frac{\sqrt{2}}{\kappa}\,.\tag{3.20}$$

Introducing harmonic expansions for the various components into (3.19) and using (3.17) to evaluate the covariant derivatives $\nabla_\pm$, one can immediately extract the equations for the harmonic components. These equations are, suppressing the iso-spin labels l, m,

$$\begin{aligned}
&\left(\partial^2-\frac{l(l+1)}{a^2}\right)(h_{ab}+\eta_{ab}h_{+-})+\eta_{ab}\frac{\sqrt{l(l+1)}}{a^2}(V_+-V_-)=-T_{ab}+\tfrac{1}{2}\eta_{ab}T_{cc}\,,\\
&\left(\partial^2-\frac{l(l+1)}{a^2}\right)h_{aa}+\frac{4}{a^2}h_{+-}+2\frac{\sqrt{l(l+1)}}{a^2}(V_+-V_-)=2T_{+-}\,,\\
&\left(\partial^2-\frac{l(l+1)}{a^2}\right)V_a+\frac{\sqrt{l(l+1)}}{a^2}(h_{a+}-h_{a-})=-J_a\,,\\
&\left(\partial^2-\frac{l(l+1)}{a^2}\right)h_{a\pm}\mp i\frac{\sqrt{2}}{a}\partial_a V_\pm\mp\frac{\sqrt{l(l+1)}}{a^2}V_a=-T_{a\pm}\,,\\
&\left(\partial^2-\frac{l(l+1)}{a^2}\right)V_\pm\pm i\frac{\sqrt{2}}{a}\partial_b h_{b\pm}\mp\frac{\sqrt{l(l+1)}}{a^2}(h_{+-}-\tfrac{1}{2}h_{bb})=-J_\pm\,,\\
&\left(\partial^2-\frac{l(l+1)-2}{a^2}\right)h_{\pm\pm}=-T_{\pm\pm}\,.
\end{aligned}\tag{3.21}$$

The solutions of these equations are to be substituted into the expression

$$\begin{aligned}
I(T,J)&=\tfrac{1}{2}\int d^6z\,\sqrt{-\bar{g}}(\tfrac{1}{2}T_{AB}h_{AB}+J_A V_A)\\
&=2\pi a^2\int d^4x\sum_{lm}[\tfrac{1}{2}\bar{T}_{ab}h_{ab}+\tfrac{1}{2}\bar{T}_{++}h_{++}+\tfrac{1}{2}\bar{T}_{--}h_{--}\\
&\qquad+\bar{T}_{a+}h_{a+}+\bar{T}_{a-}h_{a-}+\tfrac{1}{4}\bar{T}h+\bar{J}_a V_a+\bar{J}_+V_++\bar{J}_-V_-]\,.
\end{aligned}\tag{3.22}$$

Because of the reality of h_{AB} and V_A and the properties of $D^l_{\lambda m}$ it follows that the harmonic components are subject to the reality condition [9]

$$\overline{\phi^{l}_{\lambda m}(x)} = (-)^{\lambda - m} \phi^{l}_{-\lambda,-m}(x) \,. \tag{3.23}$$

This ensures the reality of $I(T, J)$.

To solve the set (3.21) and substitute the result into (3.22) is a straightforward though lengthy computation. We illustrate the approach by finding part of the dependence* of I of T_{++}. The contribution of h_{++} can be found quite easily since h_{++} decouples from the rest. Thus, from (3.21)

$$h_{++} = -\left(\partial^2 - \frac{l(l+1)-2}{a^2}\right)^{-1} T_{++} \,.$$

The contribution to I is then

$$\begin{aligned} 2\pi a^2 \int \mathrm{d}^4x \sum_{lm} [\tfrac{1}{2}\bar{T}_{++}h_{++} + \tfrac{1}{2}\bar{T}_{--}h_{--}) \\ = 2\pi a^2 \int \mathrm{d}^4x \sum_{lm} \left[-\tfrac{1}{2}\bar{T}_{++}\left(\partial^2 - \frac{l(l+1)^{-2}}{a^2}\right)^{-1} T_{++} + \text{h.c.}\right] \\ = 2\pi a^2 \int \mathrm{d}^4x \sum_{l \geq 2} \sum_{m} \bar{T}_{++}\left(-\partial^2 + \frac{l(l+1)-2}{a^2}\right)^{-1} T_{++} \,. \end{aligned}$$

To interpret this result one can Fourier analyze T_{++} obtaining

$$2\pi a^2 \int \frac{\mathrm{d}^4 p}{(2\pi)^4} \sum_{l \geq 2} \sum_{m} \frac{|\tilde{T}_{++}(p)|^2}{p^2 + [l(l+1)-2]/a^2} \,.$$

One sees that it describes the exchange of particles of mass $(l(l+1)-2)/a^2$ carrying quantum numbers appropriate to the vertex factor $\tilde{T}^{l}_{++m}(p)$, i.e. $2l+1$-dimensional multiplets of scalar particles. The overall factor $2\pi a^2$ is not significant, it can be scaled away.

It is important to make use of the conservation laws (3.5) since only the physical part of the spectrum is of interest. With the help of (3.17) one can extract from (3.5) the conservation laws for the harmonic components,

$$\begin{gathered} \partial_a T_{ab} + \frac{i}{a}\sqrt{\frac{l(l+1)}{2}}\,(T_{b+} + T_{b-}) = 0 \,, \\ \partial_a T_{a\pm} + \frac{i}{2}\sqrt{\frac{l(l+1)}{2}}\,T_{-+} + \frac{i}{a}\sqrt{\frac{(l+2)(l-1)}{2}}\,T_{\pm\pm} = \pm i\frac{\sqrt{2}}{a}J_{\pm} \,, \\ \partial_a J_a + \frac{i}{a}\sqrt{\frac{l(l+1)}{2}}\,(J_+ + J_-) = 0 \,. \end{gathered} \tag{3.24}$$

The solution of (3.21) is given in the appendix. Here we exhibit only the pole terms which result when the subsitution into (3.22) is made. The expressions are simplified by referring them to an appropriate Lorentz frame. To see the massive states it is best to use the rest frame, $p_a = (p_0, 0, 0, 0)$. For massless states we use instead the frame $p_a = (p_0, 0, 0, p_3)$. The spin of the exchanged particle is signalled by the source currents appearing in the residue.

---

* The full dependence of I on T_{++} can only be obtained by solving the complete set of equations

Taking first the massive excitations one can show that the pole terms in $I(T, J)$ arrange themselves into six towers,

$$I = \sum_{l \geq 0} (I_0^{(+)} + I_0^{(-)} + I_2) + \sum_{l \geq 1} (I_1^{(+)} + I_1^{(-)}) + \sum_{l \geq 2} I_0 , \tag{3.25}$$

where the subscripts 0, 1, 2 indicate the spin of the exchanged particles which contribute. The explicit forms are*

$$\begin{aligned}
I_0 &= \frac{1}{p^2 + M_0^2} \frac{1}{2l(l+1)} |T_{++} - T_{--}|^2 , \\
I_0^{(\pm)} &= \frac{1}{p^2 + M_{0\pm}^2} \frac{1}{4\sqrt{1+12l(l+1)}} \Bigg| \sqrt{\tfrac{4}{8}} [5 \pm \sqrt{1+12l(l+1)}]^{1/2} (\tfrac{3}{4} T_{00} + \tfrac{1}{4} T_{kk}) \\
&\quad + \left[\frac{\pm(3l(l+1)-2)\sqrt{1+12l(l+1)} - 9l(l+1) - 2}{l(l+1)-2} \right]^{1/2} T_{+-} \\
&\quad + i \left[\frac{l(l+1)}{l(l+1)-2} (5 \mp \sqrt{1+12l(l+1)}) \right]^{1/2} (J_+ + J_-) \Bigg|^2 , \\
I_1^{(\pm)} &= \frac{1}{p^2 + M_{1\pm}^2} |T_{k+} - T_{k-} \pm \sqrt{2} J_k|^2 , \\
I_2 &= \frac{1}{p^2 + M_2^2} |T_{jk} - \tfrac{1}{3} \delta_{jk} T_{ii}|^2 ,
\end{aligned} \tag{3.26}$$

where the non-zero masses are given by

$$\begin{aligned}
M_0^2 &= (l-1)(l+2)/a^2 , && l \geq 2 \\
M_{0\pm}^2 &= (2l(l+1) + 1 \pm \sqrt{1+12l(l+1)})/2a^2 , && l \geq 0 \\
M_{1\pm}^2 &= (l(l+1) \pm \sqrt{2l(l+1)})/a^2 , && l \geq 1 \\
M_2^2 &= l(l+1)/a^2 , && l \geq 0 .
\end{aligned} \tag{3.27}$$

It must be emphasized that these mass formulae have been derived by using the rest frame so that only the non-zero values belong to the spectrum. It will be observed that $M_{0-} = 0$ for $l = 0$, 1, $M_{1-} = 0$ for $l = 1$, and $M_2 = 0$ for $l = 0$. These zeroes are artifacts of the computational procedure. To find the true massless states it is necessary to repeat the computation in a more appropriate frame, $p_a =$

---

*It has been shown by Schellekens (private communication) that the residue I_0 is zero. We have confirmed this in the light cone gauge.

$(p_0, 0, 0, p_3)$. When this is done, one finds the terms

$$I(M=0)=\frac{1}{p^2}\{[\tfrac{1}{4}|T_{11}-T_{22}|^2+|T_{12}|^2+|J_1|^2+|J_2|^2]_{l=0}$$
$$+[|T_{1+}+T_{1-}-\sqrt{2}J_1|^2+|T_{2+}+T_{2-}-\sqrt{2}J_2|^2]_{l=1}\}\,, \qquad (3.28)$$

indicating a total of ten massless states: the graviton (helicities $\lambda=\pm 2$), a "photon" ($\lambda=\pm 1$), and a Yang–Mills triplet ($\lambda=\pm 1$, $l=1$). Note, in particular, the absence of zero-helicity massless states.

Finally, we observe that there are no tachyons and, since the residues all have the same sign, no negative metric states. This means that the ground state is at least perturbatively stable.

4. Zero-mode ansatz

The purpose of this section is to derive an effective four-dimensional lagrangian for the long-range sector of the six-dimensional system. That is, we shall suppose that the radius a tends to zero while discarding from the theory all those harmonic components corresponding to masses $\sim 1/a$. This is not an entirely satisfactory kind of limit: it corresponds, in fact, to letting the Maxwell gauge coupling grow without limit,

$$e=\frac{n}{2\sqrt{2}}\frac{\kappa}{a}\to\infty\,.$$

Of course, the model cannot pretend to be realistic, and with the limited set of parameters available we cannot avoid this difficulty. Our aim here is only to illustrate the procedure of "dimensional reduction" in the context of a particularly simple model.

One of the features to be brought out is the rather complicated placing of the SU(2) gauge field in the six-dimensional system. As mentioned in sect. 1, the Yang–Mills vector turns out to be a mixture of the six-dimensional metric and Maxwell fields. This is contrary to the common assumption that Yang–Mills fields are purely metrical in origin. We shall argue, however, that such mixing effects are to be expected whenever the "matter" fields which participate in the spontaneous compactification carry the appropriate quantum numbers.

Since we know from the analysis of sect. 3 that no scalar zero-modes are present we do not need to introduce any scalar fields in the zero-mode ansatz. We need to account only for the graviton and the $\mathrm{SU}(2)\times\mathrm{U}(1)$ vectors. We also know from sect. 3 that the helicity-1 triplet is a mixture of metric and Maxwell fields. In order to take account of this mixing we shall start by introducing two independent, $l=1$, 4-vectors in the ansatz. One combination will reveal itself as massive and we shall then discard it. Thus, our starting ansatz, in the notation of 1-forms*, reads

$$\begin{aligned} E^a(x,y) &= \mathrm{d}x^m E^a_m(x)\,, \\ E^\alpha(x,y) &= a\,\mathrm{d}y^\mu e^\alpha_\mu(y) - \kappa\,\mathrm{d}x^m\, W^{\hat\alpha}_m(x) D^\alpha_{\hat\alpha}(L_y) \\ &= ae^\alpha(y) - \kappa W^{\hat\alpha}(x) D^\alpha_{\hat\alpha}(L_y)\,, \qquad (4.1) \\ A(x,y) &= \mathrm{d}x^m V_m(x) + \frac{n}{2e}\left(\mathrm{d}y^\mu\, e^3_\mu(y) - \frac{\kappa}{a}\mathrm{d}x^m\, U^{\hat\alpha}_m(x) D^3_{\hat\alpha}(L_y)\right) \\ &= V(x) + \frac{n}{2e}\left(e^3(y) - \frac{\kappa}{a} U^{\hat\alpha}(x) D^3_{\hat\alpha}(L_y)\right), \end{aligned}$$

where $a = 0, 1, 2, 3$; $\alpha = +, -$; $\hat\alpha = 1, 2, 3$ and $y^\mu = (\theta, \varphi)$. At this stage the system includes the graviton field, $E^a_m(x)$, a U(1) vector, $V_m(x)$, and two SU(2) vectors, $W^{\hat\alpha}_m(x)$ and $U^{\hat\alpha}_m(x)$.

The symmetry group of the background includes rotations of the 2-sphere coupled to appropriate frame rotations and U(1) gauge transformations. The frame rotations are determined by the invariance of $e^\alpha(y)$ and the U(1) transformations by the invarince of $e^3(y)$. As discussed in sect. 3, a left translation, g, of the quotient space, SU(2)/U(1) (parametrized by L_y), induces a U(1) rotation, h, such that $gL_y = L_{y'}h$. On writing $h = \exp(\zeta Q_3)$ one finds

$$e^\pm(y') = e^\pm(y)\exp(\pm i\zeta)\,, \qquad e^3(y') = e^3(y) - \mathrm{d}\zeta\,, \qquad (4.2)$$

where $\zeta = \zeta(y, g)$. Invariance is obtained if the frames $E^\pm$ are rotated through ζ and the Maxwell field A undergoes a gauge transformation with parameter $\Lambda = n\zeta/2$. This much is enough to give invariance under x-independent left-translations. In order to get local invariance it is necessary to consider the behaviour of $W^{\hat\alpha}$ and $U^{\hat\alpha}$. Although these fields are associated with excitations on the background, their transformations include inhomogeneous terms whose contributions are needed in order to make manifest the background invariance with respect to *local* left-translations on S^2. The local version** of (4.2) is

$$\begin{aligned} e^\alpha(y') &= e^\beta(y) D^\alpha_\beta(h^{-1}) + (g^{-1}\,\mathrm{d}g)^{\hat\beta} D^\alpha_{\hat\beta}(L_y h^{-1})\,, \\ e^3(y') &= e^3(y) - \mathrm{d}\zeta + (g^{-1}\,\mathrm{d}g)^{\hat\beta} D^3_{\hat\beta}(L_y)\,, \end{aligned} \qquad (4.3)$$

where $g^{-1}\,\mathrm{d}g = \mathrm{d}x^m\, g^{-1}\partial_m g$ belongs to the algebra of SU(2). Taken in conjunction with the ansatz (4.1) these identities allow one to deduce the transformation properties of V, W, U. The group acts on E^α and A such that

$$\begin{aligned} E^\alpha(x,y) &\to E'^\alpha(x,y') = E^\beta(x,y) D^\alpha_\beta(h^{-1})\,, \\ A(x,y) &\to A'(x,y') = A(x,y) - \frac{n}{2e}\,\mathrm{d}\zeta\,. \end{aligned} \qquad (4.4)$$

---

[*] The 1-forms $e^\pm(y)$ which appear here are related to $E^\pm(y)$ defined in eq. (3.11) by $ae^\pm(y) = E^\pm(y)$.

[**] Ref. [4], appendix 3.

where A_{ab} denotes the Yang–Mills field strength made out of A and the covariant derivative ∇_a is now defined by

$$\nabla_a X_b = \partial_a X_b - \frac{\sqrt{3}}{2}\frac{\kappa}{a} A_a \times X_b \,. \tag{4.12}$$

It now appears that A_a is the true Yang–Mills vector, while X_a is massive. (It corresponds to the mass $M_{1+}(l=1)$ of eq. (3.27).) In passing, we note that the SU(2) coupling constant is given by

$$f = \frac{\sqrt{3}}{2}\frac{\kappa}{a} = \frac{\sqrt{6}}{n} e \,, \tag{4.13}$$

where e is the U(1) coupling.

This concludes our discussion of the bosonic part of the long-range sector of this model. In the next section we consider the introduction of fermions and show how fermionic zero-modes can arise.

5. Fermion spectrum

At the classical level fermion fields do not affect the ground state solution and the question of its stability. However, after establishing the properties of the classical ground state one is free to consider how it affects the propagation of fermionic excitations. In particular, one can demonstrate the presence of zero-mode solutions of the Dirac equation.

Fermion fields should be treated as scalars under the group of general co-ordinate transformations and as spinors under the tangent space group SO(1, 5). The simplest examples are 4-spinors and we shall treat only these. The relevant Clifford algebra is generated by eight-dimensional matrices, Γ_A, which satisfy

$$\begin{aligned} \tfrac{1}{2}\{\Gamma_A, \Gamma_B\} &= \eta_{AB} \\ &= \operatorname{diag}(-1, 1, 1, 1, 1, 1) \,. \end{aligned} \tag{5.1}$$

(On occasion we shall also use the combinations $\Gamma_5 \mp i\Gamma_6 = \sqrt{2}\Gamma_\pm$ appropriate to the spherical basis on the 2-sphere, i.e. $\eta_{+-} = \eta_{-+} = 1$, $\eta_{++} = \eta_{--} = 0$.) It is convenient to employ a particular realization of the algebra (5.1) which is adapted to the reduction, SO(1, 5) → SO(1, 3) × SO(2). In this realization we have

$$\begin{aligned} &\Gamma_a = \gamma_a \times \tau_1 \,, \\ &\Gamma_5 = \gamma_5 \times \tau_1 \,, \qquad a = 0, 1, 2, 3 \,, \\ &\Gamma_6 = 1 \times \tau_2 \,, \end{aligned} \tag{5.2}$$

where γ_a and γ_5 denote the usual four-dimensional matrices,

$$\tfrac{1}{2}\{\gamma_a, \gamma_b\} = \eta_{ab} \,, \qquad \gamma_5^2 = 1 \,. \tag{5.3}$$

Using (4.3) it is a simple matter to show that the form of the ansatz is maintained provided

$$\begin{aligned} U'^{\hat{\alpha}}(x) &= U^{\hat{\beta}}(x) D^{\hat{\alpha}}_{\hat{\beta}}(g^{-1}) - \frac{a}{\kappa}(g\, dg^{-1})^{\hat{\alpha}}\,, \\ W'^{\hat{\alpha}}(x) &= W^{\hat{\beta}}(x) D^{\hat{\alpha}}_{\hat{\beta}}(g^{-1}) - \frac{a}{\kappa}(g\, dg^{-1})^{\hat{\alpha}}\,. \end{aligned} \tag{4.5}$$

That is, *both* U and W transform as SU(2) gauge fields. (The U(1) vector, $V(x)$, is invariant.) It is now clear that the difference, $U - W$, transforms homogeneously and could therefore be expected to acquire a mass. This indeed happens. A straightforward computation gives the field strength 2-form, $F = \mathrm{d}A$,

$$\begin{aligned} F = {} & -\tfrac{1}{2}\, \mathrm{d}x^m \wedge \mathrm{d}x^n \{V_{mn} - \sqrt{2}(W^{\hat{\alpha}}_{mn} + \nabla_m (U - W)^{\hat{\alpha}}_n - \nabla_n (U - W)^{\hat{\alpha}}_m) D^3_{\hat{\alpha}}(L)\} \\ & + \mathrm{d}x^m \wedge E^\beta \frac{\sqrt{2}}{a} (U - W)^{\hat{\alpha}}_m D_{\hat{\alpha}}{}^{\gamma}(L) \varepsilon_{\beta\gamma 3} \\ & -\tfrac{1}{2} E^\alpha \wedge E^\beta \frac{n}{2ea^2} \varepsilon_{\alpha\beta 3}\,, \end{aligned} \tag{4.6}$$

where ∇_m denotes the SU(2) covariant derivative

$$\nabla_m (U - W)^{\hat{\alpha}}_m = \partial_m (U - W)^{\hat{\alpha}}_n - \frac{\kappa}{a} W^{\hat{\beta}}_m (U - W)^{\hat{\gamma}}_n \varepsilon_{\hat{\beta}\hat{\gamma}\hat{\alpha}}\,, \tag{4.7}$$

and W_{mn} the Yang–Mills field strength. When the components, F_{AB}, are substituted into the lagrangian and the integration over y is performed, there results the expression,

$$-\tfrac{1}{4} V^2_{ab} - \tfrac{1}{6}(W_{ab} + \nabla_a (U - W)_b - \nabla_b (U - W)_a)^2 - \frac{1}{6}\frac{4}{a^2}(U_a - W_a)^2 - \frac{n^2}{8e^2 a^4}\,. \tag{4.8}$$

On the other hand, the vector W appears in E^α and so will contribute to the six-dimensional curvature scalar, R_6. The contribution of R_6 to the four-dimensional lagrangian has been discussed in the literature [10]. It gives:

$$-\frac{1}{\kappa^2} R_4 - \tfrac{1}{6} W^2_{ab} + \frac{1}{\kappa^2 a^2}\,. \tag{4.9}$$

On combining (4.9) with (4.8) and introducing the new combinations,

$$A = \sqrt{\tfrac{1}{3}}(U + W)\,, \qquad X = \sqrt{\tfrac{1}{3}}(U - W)\,, \tag{4.10}$$

one finds,

$$\begin{aligned} & -\frac{1}{\kappa^2} R_4 - \tfrac{1}{4} V^2_{ab} - \frac{1}{8}\left[A_{ab} - (\nabla_a X_b - \nabla_b X_a) - \frac{\kappa}{a}\frac{\sqrt{3}}{2} X_a \times X_b\right]^2 \\ & \qquad - \frac{1}{8}\left[A_{ab} + (\nabla_a X_b - \nabla_b X_a) + \frac{3\kappa}{a}\frac{\sqrt{3}}{2} X_a \times X_b\right]^2 - \frac{1}{2}\frac{4}{a^2} X^2_a \\ & \quad = -\frac{1}{\kappa^2} R_4 - \tfrac{1}{4} V^2_{ab} - \tfrac{1}{4} A^2_{ab} - \tfrac{1}{4}(\nabla_a X_b - \nabla_b X_a)^2 - \frac{1}{2}\frac{4}{a^2} X^2_a + \cdots\,, \end{aligned} \tag{4.11}$$

The invariance group of the ground state involves the simultaneous action of left-translations on the 2-sphere, rotations of the E^5 and E^6 frames and a U(1) gauge transformation. In the notation of (3.13), the tangent space rotation through the angle $\zeta = \zeta(\theta, \varphi, g)$ must be accompanied by a gauge transformation with $\Lambda = +\frac{1}{2}n\zeta$ in order to preserve the form of $\bar{A}$, as discussed in sect. 4. With respect to this background symmetry the effective iso-helicities of ψ_{R} and ψ_{L} become $\frac{1}{2}-\frac{1}{2}n$ and $-\frac{1}{2}-\frac{1}{2}n$, respectively, as indicated in (5.13).

Now consider the Dirac operator (using $\tau_3\psi = \psi$),

$$\begin{aligned}\Gamma^A\nabla_A\psi &= (\gamma^a\tau_1\partial_a + \gamma^5\tau_1\nabla_5 + \tau_2\nabla_6)\psi \\ &= \tau_1(\not{\partial} + \gamma_5\nabla_5 + i\nabla_6)\psi \\ &= \tau_1\left(\not{\partial} + \frac{1+\gamma_5}{2}\sqrt{2}\nabla_- - \frac{1-\gamma_5}{2}\sqrt{2}\nabla_+\right)\psi\,.\end{aligned}$$

It appears that the mass term for the 4-spinor takes the form

$$\sqrt{2}(\bar{\psi}_{\mathrm{L}}\nabla_-\psi_{\mathrm{R}} - \bar{\psi}_{\mathrm{R}}\nabla_+\psi_{\mathrm{L}})\,. \tag{5.14}$$

To evaluate the covariant derivatives in this expression we must introduce the harmonic expansions,

$$\psi_{\mathrm{R,L}} = \sum_{l \geqslant |\lambda_{\mathrm{R,L}}|} \sqrt{2l+1}\sum_m D^l_{\lambda m}(L^{-1}_{\theta\varphi})\psi^l_{\mathrm{R,L}m}(x)\,, \tag{5.15}$$

where the relevant iso-helicities are

$$\lambda_{\mathrm{R}} = \frac{1-n}{2}\,, \qquad \lambda_{\mathrm{L}} = -\frac{1+n}{2}\,. \tag{5.16}$$

In terms of the harmonic components the mass term (5.13) becomes

$$\frac{i}{a}\sum_{lm}\sqrt{(l+\tfrac{1}{2})^2 - \tfrac{1}{4}n^2}\,(\bar{\psi}_{\mathrm{L}}\psi_{\mathrm{R}} - \bar{\psi}_{\mathrm{R}}\psi_{\mathrm{L}})\,. \tag{5.17}$$

There is evidently one massless multiplet with

$$l = -\tfrac{1}{2} + \tfrac{1}{2}|n|\,.$$

This value of l contributes to ψ_{R} but not to ψ_{L} if n is a positive integer. Conversely, if n is negative then ψ_{L} contains the massless mode*.

The coupling of the massless fermion field to the boson fields V, A and E can be derived by substituting the zero-mode ansatz into the six-dimensional Dirac lagrangian

$$i\bar{\psi}\Gamma^A E^M_A(\partial_M + B_M + ieA_M)\psi\,, \tag{5.18}$$

---

* The n zero-modes of $\Gamma_+\nabla_- + \Gamma_-\nabla_+$ associated with the magnetic charge, n, are of course predicted by the Atiyah–Singer index theorem.

The generators of SO(1, 5) are given by

$$-\tfrac{1}{2}\Gamma_{AB} = -\tfrac{1}{4}\{\Gamma_A, \Gamma_B\} \tag{5.4}$$

and they evidently commute with τ_3. The fundamental representations, 4 and $\bar{4}$, correspond to the eigenvalues +1 and −1 or τ_3, respectively.

Let $\psi(x; y)$ belong to the representation 4, i.e.

$$\tau_3\psi = \psi\,. \tag{5.5}$$

From (5.2), (5.4) we see that the SO(2) group of rotations in the 5–6 plane is generated by

$$-\tfrac{1}{2}\Gamma_{56} = -\tfrac{1}{2}i\gamma_5\tau_3\,. \tag{5.6}$$

Acting on the chiral components of ψ it gives

$$-\tfrac{1}{2}\Gamma_{56}\psi_{\mathrm{L}} = \tfrac{1}{2}i\psi_{\mathrm{L}}\,, \qquad -\tfrac{1}{2}\Gamma_{56}\psi_{\mathrm{R}} = -\tfrac{1}{2}i\psi_{\mathrm{R}}\,, \tag{5.7}$$

where ψ_{L} and ψ_{R} correspond to $\gamma_5 = -1$ and $\gamma_5 = +1$, respectively.

Now suppose that $\psi(x, y)$ carries the U(1) charge, e. Under local U(1) transformations defined by

$$A \to A - \frac{1}{e}\,\mathrm{d}\Lambda \tag{5.8}$$

we have

$$\psi \to \mathrm{e}^{i\Lambda}\psi\,, \tag{5.9}$$

where $\Lambda = \Lambda(x, y)$ is a real scalar. Derivatives of ψ which are covariant with respect to both the frame rotations, SO(1, 5), and the U(1) gauge transformations are contained in the 1-form

$$\nabla\psi = \mathrm{d}\psi + B\psi + ieA\psi\,, \tag{5.10}$$

where B denotes the spin connection 1-form

$$B = \tfrac{1}{2}\,\mathrm{d}z^M B_{M[AB]}(-\tfrac{1}{2}\Gamma_{AB})\,. \tag{5.11}$$

For the purposes of this section, where only fermion bilinears are to be retained, it will be sufficient to use background values for B and A, viz.

$$\bar{B} = -\tfrac{1}{2}\Gamma_{56}e^3\,, \qquad \bar{A} = \frac{n}{2e}\,e^3\,, \tag{5.12}$$

where e^3 is the 1-form, $-\mathrm{d}\varphi(\cos\theta - 1)$, given in sect. 3. In effect, the background covariant derivative of ψ is given by

$$\begin{aligned}\bar{\nabla}\psi &= \mathrm{d}\psi + \tfrac{1}{2}i(n + i\Gamma_{56})e^3\psi \\ &= \mathrm{d}\psi + i\left(\frac{1+\gamma_5}{2}\,\frac{n-1}{2} + \frac{1-\gamma_5}{2}\,\frac{n+1}{2}\right)e^3\psi\,.\end{aligned} \tag{5.13}$$

and integrating over y. Because of the local $SU(2)\times U(1)$ symmetry, the result must be

$$i\bar{\psi}_R\gamma^a E_a^m(\partial_m+\tfrac{1}{4}B_{m[bc]}\gamma^{bc}-fA_m^{\hat{\alpha}}I_{\hat{\alpha}}+ieV_m)\psi_R\,, \tag{5.19}$$

if ψ_R is the massless field ($n>0$). The fields ψ_R belong to the right-handed doublet of SO(1, 3) and to the n-dimensional representation of SU(2) (with generators denoted $I_{\hat{\alpha}}$). The SO(1, 3) spin connection is denoted $B_{m[bc]}$.

One may add more than one spinor field to the lagrangian: for example, consider two fields ψ, ψ' with U(1) couplings Y and Y'. From the iso-helicity formulae for ψ:

$$\lambda_R=\frac{1-nY}{2}, \qquad \lambda_L=-\left(\frac{1+nY}{2}\right),$$

and with similar formulae for ψ' with Y replaced by Y', we would obtain for the $n=1$ case, a massless iso-singlet ψ_R for $Y=1$, and a massless iso-doublet ψ'_L for $Y'=-2$. For $n=2$, with $Y=1$, $Y'=-1$, we would obtain two "iso-doublets" ψ_R and ψ'_L, both massless. For $n=2$, the formulae for the ratio of U(1) and SU(2) couplings, $\tan\theta_0=\sqrt{\frac{1}{6}}n$ (cf. eq. (4.13)) yields $\sin^2\theta=\frac{2}{5}$ – a value very close to the conventional unrenormalized value $\frac{3}{8}$. A determination of "renormalized" $\sin^2\theta$ with $\sin^2\theta_0=\frac{2}{5}$ as input in a usual GUT context would give a 3% deviation from the value deduced with $\sin^2\theta_0=\frac{3}{8}$ as input. This accidental agreement is perhaps devoid of physical significance, since we do not expect ψ_R and ψ'_L to correspond to the leptonic or quark doublets. Even so, the fact that $\sin^2\theta$ can in principle be determined by the n-quantum number (corresponding to the monopole-charge in the internal space) is an important consequence of the Kaluza–Klein picture of the influence of extra dimensions on physics.

6. Conclusion

The model discussed in this article, six-dimensional Einstein–Maxwell theory (with fermions), is not intended as a realistic physical theory. We have analyzed it in some detail because it seems to represent the simplest possible version of a Kaluza–Klein mechanism which yet contains some of the features to be hoped for in more elaborate systems: compactification arises as a consequence of dynamics, the internal space has a non-abelian symmetry, massless chiral fermions are available, and the zero-mode sector includes no scalars. This model suffers the defects which are generally expected as well: it is not renormalizable, and the gauge couplings grow unacceptably large in the very limit which is needed to drive away the towers of massive modes. Another defect is the need for an "unnatural" adjustment of parameters in the parent lagrangian in order to flatten the four-dimensional spacetime. The presence of such adjustable parameters is both a defect and a virtue. One would like to be able to start from a less arbitrary theory, a theory in which the "matter" content and its couplings are fixed. For this reason, the many-dimensional supergravities are attractive. Up to now, however, such theories have not yielded any physically acceptable ground state solution. Either the internal space is flat [10], with abelian symmetry or, if it is curved, it generates four-dimensional curvature of the same magnitude [5]. There are too few adjustable parameters. One hopes that this is a temporary impasse.

Finally, we believe that our approach to the phenomenon of dimensional reduction, via the computation of spectra, illuminates an important matter. This is the question of what four-dimensional fields should be included in the zero-mode ansatz, and how they should be arranged. Without first going through the spectral analysis one cannot foresee the correct form of the ansatz. Although the gauge vectors are guaranteed by the background symmetry, their placement in the ansatz has to be found. The very presence or absence of massless scalars seems to be model dependent (except when the background is supersymmetric, in which case they are certain to be present). Quite generally, the underlying mechanism whereby "matter" fields generate curvature on the internal space would seem to imply a non-trivial mixing between matter and metrical field components. This mixing has to be correctly represented in the zero-mode ansatz.

We have benefited from many stimulating and useful discussions with E. Sezgin.

Appendix

The solution to the linear eq. (3.19) which was used to obtain the pole structures (3.26)–(3.28) is listed here for reference

$$h_{ab}-\tfrac{1}{4}\eta_{ab}h_{cc}=\frac{1}{p^2+M_2^2}(T_{ab}-\tfrac{1}{4}\eta_{ab}T_{cc})\,,$$

$$\begin{aligned}h_{aa}=\frac{1}{(p^2+M_2^2)(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\Big[&-2(p^2+M_{1-}^2)(p^2+M_{1+}^2)T_{+-}\\&+\frac{p^2}{a^2}T_{aa}+\frac{2}{a}M_2(p^2+M_0^2)(J_+-J_-)+\frac{2M_0M_2}{a^2}(T_{++}+T_{--})\Big]\,,\end{aligned}$$

$$\begin{aligned}h_{+-}=\frac{1}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\Big[&\tfrac{1}{4}(p^2+M_2^2)(2T_{+-}-T_{aa})\\&-\frac{3}{2}\frac{M_2}{a}(J_+-J_-)-\frac{3}{2}\frac{M_0M_2}{a^2(p^2+M_0^2)}(T_{++}+T_{--})\Big]\,,\end{aligned}$$

$$h_{++}=\frac{1}{p^2+M_0^2}T_{++}\,,$$

$$\begin{aligned}
V_a = {} & \frac{\sqrt{2}M_2 p_a}{a^2(p^2+M_2^2)(p^2+M_{1+}^2)(p^2+M_{1-}^2)}\left[\frac{-M_0}{a(p^2+M_0^2)}(T_{++}-T_{--})-(J_++J_-)\right] \\
& +\frac{1}{(p^2+M_{1+}^2)(p^2+M_{1-}^2)}\left[\frac{M_2}{a}(T_{a+}-T_{a-})+(p^2+M_2^2)J_a\right],
\end{aligned}$$

$$\begin{aligned}
2V_- = {} & \frac{1}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\left[\frac{M_2}{a}(T_{+-}-\tfrac{1}{2}T_{aa})-\left(p^2+M_2^2+\frac{1}{a^2}\right)(J_+-J_-)\right. \\
& +\frac{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}{p^2+M_2^2}(J_++J_-)-\frac{M_0}{a}\,\frac{p^2+M_2^2+1/a^2}{p^2+M_0^2}(T_{++}+T_{--}) \\
& \left.+\frac{M_0}{a}\,\frac{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}{(p^2+M_0^2)(p^2+M_2^2)}(T_{++}-T_{--})\right],
\end{aligned}$$

$$\begin{aligned}
2h_{a-} = {} & \frac{-\sqrt{2}M_2 p_a}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\left[T_{+-}+\frac{1}{a^2(p^2+M_2^2)}T_{bb}\right] \\
& +\frac{1}{a(p^2+M_{1+}^2)(p^2+M_{1-}^2)}[-2M_2J_a+\sqrt{2}p_a(J_++J_-)] \\
& +\frac{\sqrt{2}M_0 p_a}{a^2(p^2+M_{1+}^2)(p^2+M_{1-}^2)(p^2+M_0^2)}(T_{++}-T_{--}) \\
& -\frac{M_0}{\sqrt{2}}\,\frac{(p^2+M_2^2+1/a^2)p_a}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)(p^2+M_0^2)}(T_{++}+T_{--}) \\
& +\left[\left(\frac{-(p^2+M_2^2)}{(p^2+M_{1+}^2)(p^2+M_{1-}^2)}+\frac{1}{p^2+M_2^2}\right)\eta_{ab}\right. \\
& \left.-\frac{(p^2+M_2^2+1/a^2)p_ap_b}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)(p^2+M_2^2)}\right]T_{b+} \\
& +\left[\left(\frac{(p^2+M_2^2)}{(p^2+M_{1+}^2)(p^2+M_{1-}^2)}+\frac{1}{p^2+M_2^2}\right)\eta_{ab}\right. \\
& \left.-\frac{(p^2+M_2^2+1/a^2)p_ap_b}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)(p^2+M_2^2)}\right]T_{b-}\,.
\end{aligned}$$

References

[1] Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, Nucl. Phys. B127 (1977) 57
[2] L. Palla, Tokyo Conf. Report, (1978) p. 629
[3] S. Randjbar-Daemi and R. Percacci, Phys. Lett. 117B (1982) 41
[3a] G. Chapline and R. Slansky, Los Alamos preprint LA 82-1076
[4] A. Salam and J. Strathdee, Ann. of Phys. 141 (1982) 316
[5] P.G.O. Freund and M. Rubin, Phys. Lett. 97B (1980) 233
[6] E. Sezgin, in preparation
[7] L.F. Abbot and S. Deser, Nucl. Phys. B195 (1982) 76
[8] P. van Nieuwenhuizen and S. Deser, Phys. Rev. D10 (1974) 401
[9] N.Ya. Vilenkin, Special functions and the theory of group representations, Providence, R.I. (1968)
[10] R. Percacci and S. Randjbar-Daemi, Kaluza–Klein theories on bundles with homogeneous fibres-I, ICTP, Trieste, preprint IC/82/18, J. Math. Phys. to appear
[11] J. Scherk and J.H. Schwarz, Nucl. Phys. B153 (1979) 61

QUASI-RIEMANNIAN THEORIES OF GRAVITATION IN MORE THAN FOUR DIMENSIONS

Steven WEINBERG[1]
Theory Group, Physics Department, University of Texas, Austin, TX 78712, USA

Received 16 January 1984

We explore the possibility that in $N > 4$ spacetime dimensions, gravitation is described by a theory based on a modified Principle of Equivalence, according to which the field equations in locally inertial frames obey a "tangent space" group G_T other than the Lorentz group $O(N-1, 1)$. It is found that such theories are severely restricted by the requirement that G_T have a subgroup which acts on four inertial coordinates as $O(3, 1)$ and leaves the others invariant. Theories of this sort can have chiral fermions in four dimensions even without non-gravitational gauge fields.

In the recent upsurge of work on higher-dimensional field theories, it has generally been assumed that gravitation is to be described in $4 + M$ dimensions by essentially the same Einstein theory that applies in 4 dimensions – perhaps with different field equations, but still within the same framework of riemannian geometry as for ordinary general relativity. Yet there are equally plausible alternatives. General relativity can be derived from a Principle of Equivalence: at any point X in an arbitrary gravitational field, we may construct a set of locally inertial coordinates $\xi_X{}^s$, in which the effects of gravitation are absent. To this, one adds an additional proviso, that the field equations when written in locally inertial coordinates are explicitly invariant under linear Lorentz transformations, and one can then derive the existence of a metric and the whole apparatus of riemannian geometry[‡1]. But why the Lorentz group? Of course, there is little doubt that Lorentz invariance is the correct symmetry in locally inertial frames in four dimensions, but how do we know that in $4 + M$ dimensions, we should just replace $SO(3, 1)$ with $SO(3 + M, 1)$?

This question gains urgency from the fact that conventional extra-dimensional "Kaluza–Klein" theories are having troubles. In order to get light fermions in chiral representations of the low-energy gauge group, it is necessary to add extra gauge fields, and in topologically non-trivial vacuum configurations [2]. These extra gauge fields cannot be derived from gravity in a conventional Kaluza–Klein theory of even higher dimensionality without reintroducing the problem of the light chiral fermions, so the hope of a purely gravitational theory of all interactions seems to be disappointed. One of the purposes of the present work is to circumvent this problem.

We shall adopt the Principle of Equivalence in $N = 4 + M$ dimensions, so that the neighborhood of any point X is described by a set of locally inertial coordinates $\xi_X{}^s$, with $s = 1, 2, ..., N$. However, the field equations are here assumed to be invariant under transformations

$$\xi_X{}^s \to \gamma^s{}_t(X)\xi_X{}^t + a^s(X) , \qquad (1)$$

with the $\gamma^s{}_t$ forming some unknown *tangent-space group* G_T of $N \times N$ real matrices. When the field equations are rewritten in terms of general lab coordinates x^σ, we encounter partial derivatives: a *vielbein*

$$e^s{}_\sigma(X) \equiv (\partial \xi_X{}^s/\partial x^\sigma)_{x=X} , \qquad (2)$$

and an *affine connection*

$$\Gamma^\sigma{}_{\tau\eta}(X) e^s{}_\sigma(X) \equiv (\partial^2 \xi_X{}^s/\partial x^\tau \partial x^\eta)_{x=X} . \qquad (3)$$

---

[1] Research supported in part by the Robert A. Welch Foundation and under NSF grant PHY 8304629.

[‡1] This is the point of view adopted for instance in ref. [1].

It makes little sense to suppose that a set of coordinates is locally inertial at just one point, so we add the proviso that $\xi_X{}^s$ is locally inertial at $X + \mathrm{d}X$ to at least first order of $\mathrm{d}X$, and is therefore related to $\xi_{X+\mathrm{d}X}{}^s$ by an infinitesimal $\mathrm{G_T}$-transformation [‡2]. That is,

$$\partial \xi_X{}^s(x)/\partial X^\sigma = [\omega_\sigma(X)]^s{}_t \xi_X{}^t(x) + a^s{}_\sigma(X)\,, \tag{4}$$

where $\omega_\sigma(X)$ is an $N \times N$ matrix in the Lie algebra of $\mathrm{G_T}$, known as the *spin connection.* From (2)–(4), one obtains the fundamental relation between the vielbein and connections

$$\partial e^s{}_\sigma(X)/\partial X^\tau = [\omega_\tau(X)]^s{}_t e^t{}_\sigma(X) + \Gamma^\rho{}_{\sigma\tau}(X) e^s{}_\rho(X)\,. \tag{5}$$

In the usual derivation of this result, ω and Γ are introduced as connections that allow the construction of $\mathrm{G_T}$-covariant and generally covariant derivatives, respectively, and e is the matrix that allows us to convert coordinate tensors to $\mathrm{G_T}$-tensors; eq. (5) then follows from the requirement that covariant differentiation has the same effect in either basis. The treatment here differs in that (3) requires $\Gamma^\rho{}_{\sigma\tau}$ to be symmetric in σ and τ. This absence of torsion is important not only philosophically, as part of the Equivalence Principle, but also practically: as we shall see, depending on the size of $\mathrm{G_T}$, the absence of torsion allows some or all of the components of Γ and ω to be calculated from the vielbein and its derivatives, while with arbitrary torsion all components of ω and some or all of Γ would be independent dynamical variables.

Riemannian geometry is seen here as just the special case for which the inertial coordinates are in the defining representation of $\mathrm{G} = \mathrm{SO}(N)$ [‡3]; other cases with different representations or different groups will be called *quasi-riemannian* geometries. In riemannian geometry there is an invariant $\mathrm{G_T}$-tensor η_{st} with $\eta_{st} = \eta_{ts}$ and $\mathrm{Det}\,\eta \neq 0$, and a corresponding covariantly constant coordinate tensor, the metric $g_{\sigma\tau} \equiv e^s{}_\sigma e^t{}_\tau \eta_{st}$. In quasi-riemannian geometry there may be invariant $\mathrm{G_T}$-tensors of various types and ranks corresponding to various sorts of "quasi-metric" coordinate tensors, whose covariant derivatives all vanish according to eq. (5). For instance if the inertial coordinates are the real and imaginary parts of the defining representation of $\mathrm{G_T} = \mathrm{U}(N/2)$ (with N even) there is in addition to η_{st} an antisymmetric η'_{st}, and hence an antisymmetric covariantly constant quasi-metric $g'_{\sigma\tau} = e^s{}_\sigma e^t{}_\tau \eta'_{st}$ as well as the symmetric metric $g_{\sigma\tau}$. This is Kähler geometry. For $\mathrm{G_T} = \mathrm{Sp}(N, R)$ (even N), there is an antisymmetric $g'_{\sigma\tau}$ but no symmetric $g_{\sigma\tau}$. For $\mathrm{G_T} = \mathrm{SL}(N, R)$ there is only a tensor density $\mathcal{D} \neq \mathrm{Det}\, e$, while for $\mathrm{G_T} = \mathrm{GL}(N, R)$ there are no η's or g's. Some non-semi-simple subgroups of $\mathrm{GL}(N, R)$ cannot be characterized by η's at all [‡4]. It seems generally more convenient to characterize the various quasi-riemannian spaces according to the group $\mathrm{G_T}$ and the $\mathrm{G_T}$-transformations of the inertial coordinates, rather than by giving a menu of metric and quasi-metric tensors.

It is important in physical applications to know which of the vielbein and conncetion components are independent dynamical variables. Eq. (5) can always be used to calculate the spin connection for a given vielbein and affine connection, but in the absence of torsion it generally also provides constraints on the relation between Γ and $\partial e/\partial x$, or on $\partial e/\partial x$ alone. In riemannian geometry it is well known that (5) determines Γ and ω uniquely in terms of $\partial e/\partial x$, but there are no constraints on the $\partial e/\partial x$ alone. More generally eq. (5) cannot determine Γ and ω uniquely in terms of $\partial e/\partial x$ unless the dimensionality D_T of $\mathrm{G_T}$ satisfies $D_\mathrm{T} \leqslant N(N-1)/2$, and it always yields constraints on $\partial e/\partial x$ unless $D_\mathrm{T} \geqslant N(N-1)/2$. For instance in the Kähler case $\mathrm{G_T} = \mathrm{U}(N/2)$, where $D_\mathrm{T} < N(N-1)/2$, Γ is given in terms of the metric $g_{\sigma\tau}$ just as in riemannian geometry, but $\partial e/\partial x$ is subject to a constraint, that the quasi metric $g'_{\sigma\tau}$ is also covariant constant. In the case $\mathrm{G_T} = \mathrm{SL}(N, R)$, where $D_\mathrm{T} > N(N-1)/2$, all components of Γ are independent variables except $\Gamma^\rho{}_{\rho\sigma} = \partial_\sigma \mathcal{D}$, and there are no constraints on $\partial e/\partial x$. On the other hand, for both the Galileo group [with $D_\mathrm{T} = N(N-1)/2$] and $\mathrm{G_T} = \mathrm{Sp}(N, R)$ [with $D_\mathrm{T} > N(N-1)/2$] there are components of Γ independent of $\partial e/\partial x$, and not all components of $\partial e/\partial x$ are independent.

We shall be much concerned here with *isometries:* combined coordinate transformations $x^\sigma \to x'^\sigma$ and inertial frame $\mathrm{G_T}$-transformations (1), that leave the vielbein and connections form-invariant. In the infinitesimal case, these transformations are specified by a *Killing vector* $K^\sigma(x)$ and a *Killing angle* $\theta(x)$ (a matrix in the Lie algebra of $\mathrm{G_T}$):

[‡2] See for instance section 3.3 of ref. [1].

[‡3] For brevity, here and elsewhere we use $\mathrm{SO}(N)$ to denote not only the true orthogonal group, but also all the pseudo orthogonal groups $\mathrm{SO}(N_1, N_2)$ with $N_1 + N_2 = N$.

[‡4] Henneaux [3] has given an example of two different semi-simple groups that have precisely the same invariant tensors.

$$x'^\sigma = x^\sigma + \epsilon K^\sigma(x)\,, \quad \gamma(x) = 1 + \epsilon\theta(x)\,. \tag{6}$$

The conditions of form-invariance for the vielbein and affine connection are, respectively,

$$K^\tau{}_{;\sigma} = [e^{-1}(\theta - \omega_\eta K^\eta)e]^\tau{}_\sigma\,, \tag{7}$$

and

$$K^\rho{}_{;\sigma;\tau} = R^\rho{}_{\sigma\eta\tau}K^\eta \tag{8}$$

while that for ω follows from these. (Here $R^\rho{}_{\tau\sigma\eta}$ is the Riemann–Christoffel curvature tensor. given by the usual formula [1] in terms of $\Gamma^\rho{}_{\tau\sigma}$.) Eqs. (7) and (8) have the important consequence that both $K^\tau(x)$ and $\theta(x)$ are determined everywhere by the value and first derivatives of the Killing vector at any one point. Also, eq. (7) tells us that the isotropies at a point x_0 [isometries that leave x_0 fixed] form a subgroup $\mathrm{H}(x_0)$ of $e^{-1}(x_0)\mathrm{G_T}e(x_0)$. [These conclusions apply also with torsion, though (7) and (8) are different.] It follows that the dimensionality D_I of the isometry group $\mathrm{G_I}$ is limited by $D_\mathrm{I} \leqslant N + D_\mathrm{T}$. [This is a generalization of the result in riemannian geometry, that $D_\mathrm{I} \leqslant N(N+1)/2$.] Homogeneous spaces are those on which $\mathrm{G_I}$ acts transitively, or equivalently, for which $D_\mathrm{I} = N + D_\mathrm{H}$. Maximally symmetric spaces are those homogeneous spaces for which the isotropy groups ate isomorphic to $\mathrm{G_T}$, or equivalently, those spaces with $D_\mathrm{I} = N + D_\mathrm{T}$. Homogeneous spaces can be identified as the coset spaces $\mathrm{G_I/H}$; maximally symmetric spaces are $\mathrm{G_I/G_T}$. It can be shown that the vielbein for a homogeneous space $\mathrm{G_I/H}$ is uniquely determined by $\mathrm{G_I}$ and H, up to a choice of lab coordinates, a local $\mathrm{G_T}$-transformation, and a global $\mathrm{G_T}$-automorphism.

Now we come to dimensional reduction. Without attempting here to construct a dynamical framework for compactification, we shall simply assume that there arises a stable vacuum solution for which the vielbein and connections obey an isometry $\mathrm{G_I} = \mathrm{P_4} \times \mathrm{C}_M$: $\mathrm{P_4}$ is four-dimensional Poincaré invariance, acting on the first 4 coordinates z^μ ($\mu = 1, 2, 3, 0$), and C_M is some compact group acting (for simplicity, transtively) on the remaining M coordintes y^α ($\alpha = 1, 2, ..., M$).

This isometry has important implications for the structure of the tangent space group. Note that the isotropy subgroup $\mathrm{H}(z_0, y_0) \subset \mathrm{G_I}$ contains a further subgroup, consisting of homogeneous Lorentz transformations on the z^μ coordinates, under which the y^α coordinates transform as scalars. As remarked earlier, the isotropy group of any space is always (up to a similarity transformation) a subgroup of the tangent-space group $\mathrm{G_T}$, so here $\mathrm{G_T}$ must contain a subgroup that in a suitable basis acts on the first four inertial coordinates $\xi_x{}^m$ ($m = 1, 2, 3, 0$) as a homogeneous Lorentz transformation, and leaves the remaining M inertial coordinates $\xi_x{}^a$ ($a = 1, 2, ..., M$) invariant. This will be referred to as the *Lorentz subgroup condition.*

What sort of tangent-space group can satisfy this condition? Of course, a great variety of groups of $N \times N$ real matrices have SO(3, 1) subgroups. The restrictive part of this condition is that under SO(3, 1) the inertial coordinates split up into a *single* four-vector plus $N - 4$ scalars. [for instance, the group SU(3, 1) contains a subgroup SO(3, 1), but the defining representation of SU(3, 1) consists of *two* real SO(3, 1) four-vectors.] One candidate for $\mathrm{G_T}$ that obviously does satisfy the Lorentz subgroup condition is $\mathrm{SO}(L_1\; L_2) \times \mathrm{G'_T}$, where $\mathrm{SO}(L_1, L_2)$ acts as a homogeneous Lorentz transformation on the first $L_1 + L_2$ inertial coordinates (with $L_1 \geqslant 3, L_2 \geqslant 1, L_1 + L_2 \leqslant N$), and $\mathrm{G'_T}$ is some arbitrary group of real matrices acting on the remaining $N - L_1 - L_2$ inertial coordinates. Another obvious candidate for $\mathrm{G_T}$ is obtained by replacing $\mathrm{SO}(L_1, L_2)$ with $\mathrm{SL}(L_1 + L_2, R)$. Are there any others? It can be shown [4] that these are the only semi-simple groups that satisfy the Lorentz subgroup condition. Also, these are the only possibilities for the semi-simple part of any tangent-space group.

We will pass over the possibility of a non-semi-simple tangent space group as an unattractive complication, and we reject $\mathrm{SL}(L, R) \times \mathrm{G'_T}$ because we want to include spinors in the theory. Also, causality presumably rules out $\mathrm{SO}(L_1, L_2) \times \mathrm{G'_T}$ with $L_2 > 1$. This leaves us with a tangent space group of the form

$$\mathrm{G_T} = \mathrm{SO}(L-1, 1) \times \mathrm{G'_T} \quad (4 \leqslant L \leqslant N)\,. \tag{9}$$

It can be shown [4] that by a suitable local $\mathrm{G_T}$-transformation (9), it is possible to put the vacuum vielbein in the standard form

$$e_0{}^m{}_\mu = \delta^m{}_\mu\,, \quad e_0{}^m{}_\alpha = e^a{}_{0\mu} = 0\,, \quad e^a{}_{0\alpha} = e^a{}_{0\alpha}(y)\,, \tag{10}$$

with $e^a{}_{0\alpha}(y)$ form-invariant under C_M. This form of $e_0{}^m{}_\mu$ will allow us to use μ, ν, etc., to denote spacetime components in both locally inertial and lab frames, but we continus to use α, β, ... and a, b, ... to distinguish lab and inertial frame bases for internal components.

The tangent space group (9) is small enough to provide useful constraints on the connections. It is easy to show [4] that the inertial frame components $\Gamma^s{}_{tu}$ of the affine connection can all be expressed in terms of the vielbein and its first derivatives, except perhaps (depending on G'_T) for the components with $s, t, u > L$, for which $\Gamma^s{}_{tu}$ is an $SO(L-1, 1)$ scalar. Also, as usual the spin connection is completely determined by the affine connection and vielbein and its derivatives. Hence the only independent perturbations to the vacuum vielbein and connections that appear in 4 dimensions as fields of non-zero spin are the submatrices $\delta e^\mu{}_\nu$, $\delta e^\mu{}_\alpha$, and $\delta e^a{}_\mu$ of the vielbein perturbation $\delta e^s{}_\sigma$.

All of these perturbations can be written as sums over representations of C_M, with four-vector or tensor coefficients depending on z^μ. But which of these are *massless* spin 1 fields? This is usually decided by evaluating the perturbations to the $(4+N)$-dimensional gravitational action, but we can also identify the massless spin 1 modes without knowing anything about the action. The point is that we expect some vector coefficient $V^\mu(z)$ in the expansion of δe to appear as a massless vector field in 4 dimensions if and only if (1) there is a combined general coordinate and G_T-transformation which subjects $V_\mu(z)$ to a gauge transformation, and (2) there is no such transformation that eliminates $V_\mu(z)$ altogether. On this basis, it can be shown [4] that massless spin 1 fields $V_\mu(z)$ appear in 4 dimensions as the coefficients of the Killing vector terms in $\delta e^a{}_\mu$:

$$\delta e^a{}_\mu(z,y) = \sum_i e^a{}_{0\alpha}(y) K^\alpha{}_i(y) V_{i\mu}(z) + \dots , \tag{11}$$

just as in riemannian geometry. That is, the gauge group seen in 4 dimensions is as usual the same as the isometry group C_M.

Finally we consider the problem of the chiral fermions. We will limit our discussion here to fermion fields that belong to the spin $\frac{1}{2}$ representation of the $SO(L-1, 1)$ part of the tangent space group (9). The question is whether in 4 dimensions at ordinary energies we will encounter Dirac particles of zero mass (aside from ordinary-energy symmetry breaking) whose $\gamma_5 = +1$ and $\gamma_5 = -1$ parts transform as different representations of whatever symmetry group survives down to ordinary energies, as seems to be required by experiment. For odd L (and arbitrary total dimensionality N) this is impossible; there is a single spin $\frac{1}{2}$ representation of $SO(L-1, 1)$, and it contains $\gamma_5 = +1$ and $\gamma_5 = -1$ parts with the same transformation under $SO(L-4)$ and under any additional gauge or global symmetries we may add to the theory, and hence with the same transformation under any ordinary-energy symmetries. For even L chiral fermions have a chance; there are two spin $\frac{1}{2}$ representations of $SO(L-1, 1)$ and for any one of then the $\gamma_5 = +1$ and $\gamma_5 = -1$ parts have different transformations under the $SO(L-4)$ part of the tangent space group, and hence might transform differently under the ordinary-energy symmetry group.

It is known [2,5] that chiral fermions can arise in this way in riemannian theories, provided we add suitable gauge fields in topologically non-trivial vacuum configurations. Here we ask whether in quasi-riemannian theories it is possible to find chiral fermions when we do *not* add non-geometric symmetries, so that the gauge group seen in 4 dimensions is just the isometry group C of the compact homogeneous manifold.

To answer this, it is not necessary to look for solutions of Dirac-like equations in N dimensions; as Witten [2] has emphasized, we can identify any chiral fermions by examining the harmonic expansion of the spinor field. Consider a function ψ of spacetime coordinates z^μ and internal coordinates y^α that transforms according to some definite representation $\bar{D}_T$ of the internal part $\bar{G}_T = SO(L-4) \times G'_T$ of the tangent-space group. As explained by Salam and Strathdee [6] such a ψ can be expanded in a series of functions of y^α that transform under combined y-coordinate transformations and G_T-transformations according to irreducible representations $D_C{}^r$ of C; the coefficients in this expansion are z-dependent fields that are seen in 4 dimensions as belonging to representations $D_C{}^r$ of the gauge group C. Specifically, for a homogeneous compact manifold C/H, a field ψ belonging to the representation $\bar{D}_T$ of $\bar{G}_T$ will transform under the isotropy group $H \subset C$ according to a sum of irreducible representations $D_H{}^i$; the representation $D_C{}^r$ will occur in the harmonic expansion of ψ as many times as any of the representations $D_H{}^i$ of H ⊂ C occurs in $D_C{}^r$. If a particular complex irreducible representation $D_C{}^r$ occurs $n_\pm{}^r$ times in the harmonic expansion of the $\gamma_5 = \pm 1$ parts of the spin$\frac{1}{2}$ field, and if (say) $n_+ > n_-$, then in 4 dimensions we will have $n_+ - n_-$ unmatched massless chiral fermions with $\gamma_5 = +1$ in the representation $D_C{}^r$ of the gauge group.

Here is an example. We consider a seven dimensional space, but with tangent space group taken not as SO(6, 1), but instead SO(5, 1). The internal part of the tangent space group is then simply $\bar{G}_T = SO(2)$; that is, $\bar{G}_T$ rotates ξ^5 and ξ^6 as a two-vector, and leaves ξ^7 invariant. We suppose that the compact manifold is maximally symmetric, so we take its isotropy group H as SO(2), and the isometry group C_3 as the (3 + 1)-dimensional group SO(3) × SO(2). However, to make life interesting we do not identify the two SO(2)'s; rather, the generator of H is taken more generally as

$$t_H = t_3 + kt_0 \tag{12}$$

where t_0 is the generator of the SO(2) part of C_3; t_3 is one of the SO(3) generators; and k is some constant. The parts of the spinor field with $\gamma_5 = \pm 1$ are here in representations of $\bar{G}_T = H$ with $t_H = \pm\frac{1}{2}$. Now, representations of C_3 are labelled by the SO(3) "isospin" j and the SO(2) "hypercharge" $t_0 = q$, and as seen immediately from (12), the harmonic expansion of the $\gamma_5 = \pm 1$ parts of the ψ will contain one each of all the representations (j, q) with

$$j = |\pm\tfrac{1}{2} - kq|\,,\ |\pm\tfrac{1}{2} - kq| + 1, \ldots .$$

Thus in particular kq must be a positive or negative integer or half-integer. There is a complete matching of the representations for $\gamma_5 = +1$ and $\gamma_5 = -1$, except that for $\gamma_5 = +1$ there is one extra with $j = kq - \frac{1}{2}$ for each q with $2kq$ a positive definite integer, while for $\gamma_5 = -1$ there is one extra with $j = |kq| - \frac{1}{2}$ for each q with $2kq$ a negative definite integer. These appear in 4 dimensions as an infinite tower of massless chiral fermions. This conclusion has been confirmed by direct solution of the zero-mass Dirac equation on the manifold SO(3) × SO(2)/SO(2) [3].

The initiate will recognize the similarity between this model and that of Randjbar-Daemi, Salam, and Strathdee [5] (RSS). They work in a six-dimensional riemannian space with tangent space group SO(5, 1), with 2 dimensions forming a compact manifold S^2 = SO(3)/SO(2), and add a U(1) gauge group, with vacuum gauge field given by a monopole configuration $A_\theta = 0, A_\phi = k(\cos\theta \pm 1)$ on S^2. In comparison, the vielbein of our three-dimensional SO(3) × SO(2)/SO(2) manifold has non-vanishing components $e^5{}_\theta = +1$, $e^6{}_\phi = \sin\theta$ (as for S^2); $e^7{}_\psi = 1$ (as for S^1); and $e^7{}_\phi = k\cos\theta$, like a monopole vector potential. The RSS model with a spinor of charge q contains one chiral fermion with $j = |kq| - \frac{1}{2}$ and γ_5 equal to the sign of kq. (Witten's demonstration [2] of this result provided the prototype for our analysis here of the seven-dimensional model.) In contrast, in our model we have an extra dimension instead of an extra gauge field, and in consequence we get an infinite number of massless fermions with all integer or half-integer values of $|kq|$.

Phenomenologically, an infinite number of massless chiral fermions is about as bad as none at all. However our model was chosen purely as a simple illustration[‡5] There are more complicated quasi-riemannian theories of this sort with finite numbers of chiral fermions [7,2].

It remains to be seen whether realistic theories can be constructed along these lines;

Acknowledgement. I am grateful for valuable discussions at various stages in this work with L. Alvarez-Gaumé, P. Candelas, M. Duff, P. Freund,, P. Ginsparg, M. Rubin, C. Wetterich, J. Wheeler and E. Witten.

References

- [1] S. Weinberg, Gravitation and cosmology (Wiley, New York, 1972).
- [2] E. Witten, Princeton preprint (October 1983), and earlier references quoted therein.
- [3] M. Henneaux, private communication.
- [4] S. Weinberg, paper in preparation.
- [5] S. Randjbar-Daemi, A. Salam and J. Strathdee, Nucl. Phys. B214 (1983) 491.
- [6] A. Salam and J. Strathdee, Ann. Phys. (NY) 131 (1982) 316.
- [7] E. Witten, private communication.

[‡5] One other unsatisfactory aspect of this illustrative model is the need for torsion; the internal part SO(2) of the tangent space group is so small that without torsion the constraints on the vielbein would not allow the construction of an SO(3) × SO(2)/SO(2) manifold with $k \neq 0$. Fortunately this is not a general problem.

CURLING UP TWO SPATIAL DIMENSIONS WITH SU(1,1)/U(1)

Murray GELL-MANN [1]
California Institute of Technology, Pasadena, CA 91125, USA

and

Barton ZWIEBACH [2]
Physics Department, University of California, Berkeley, CA 94720, USA

Received 19 July 1984

It is seen that a nonlinear sigma model based on the noncompact coset space SU(1,1)/U(1) can curl up two spatial dimensions into a topologically noncompact surface of finite area with a compact U(1) isometry group. This mechanism can be used for several higher-dimensional supergravity theories. In particular, chiral $N = 2, D = 10$ supergravity would reduce to an $N = 1, D = 8$ theory in which the masslessness of fermions does not depend only on supersymmetry. Further reduction to four dimensions is possible.

Nonlinear sigma models based on noncompact spaces appear in the context of supergravity theories with $D \leqslant 10$. The simplest nontrivial example of such a noncompact space is the two-dimensional hyperbolic coset space $H_2 = SU(1,1)/U(1)$ first recognized in $N= 4$ supergravity in four dimensions as describing the dynamics of a complex scalar field [1]. The sigma model in chiral $N = 2, D = 10$ supergravity is also based on H_2 [2–4]. This theory is of particular interest because it is the zero-mass sector of the superstring theory that might be completely finite [5].

We report here on the use of the H_2 scalar dynamics to curl up two spatial dimensions. These two dimensions then form a surface, the "tear drop", topologically equivalent to H_2 or to a two-sphere with a point removed. Although H_2 and the tear drop are topologically equivalent noncompact manifolds, their metrics are different. The tear drop has finite area, while the area of H_2 is infinite. For the chiral $N = 2, D = 10$ theory our result gives an $N = 1, D = 8$ effective theory without inducing a cosmological term at the classical level.

It is known that scalar fields described by a compact sigma model can produce spontaneous compactification [6,7] such that the extra dimensions form a space of the same shape as the space of the scalar fields. For the physically interesting case of $D > 4$ supergravities, however, all known sigma model spaces are noncompact and negatively curved, and the original ansatz for compactification [6,7] does not work. One would need to add a bare negative (anti-de Sitter type) cosmological term to the action [7], but such an addition is forbidden by supersymmetry. Furthermore, the spatial dimensions affected would form a noncompact space of infinite volume or area. The solution described here circumvents these difficulties and yields a noncompact surface of finite area.

The shape of that surface turns out to depend critically on the magnitude of the scalar self-coupling, and therefore we concentrate on a single case, that of $N = 2, D = 10$ chiral supergravity. It seems reasonable to assume that our mechanism will work for any other supergravity with an H_2 sigma model. Examples are given at the end of this letter.

---

[1] Work supported in part by the US Department of Energy under contract No. DE-AC03-81-ER40050 and by the Fleischmann Foundation.

[2] Work supported by the National Science Foundation under Research Grant No. PHY-82-03424.

The coupled equations of motion for gravitation and scalar fields in chiral $N = 2, D = 10$ supergravity are the following [3]:

$$R_{\mu\nu} = \tfrac{1}{2} h_{ij}(\phi)\partial_\mu \phi^i \partial_\nu \phi^j \,, \tag{1}$$

$$(2/|g|^{1/2})\partial_\mu(|g|^{1/2} g^{\mu\nu} h_{ij}\partial_\nu \phi^j) = g^{\mu\nu}(\partial h_{pq}/\partial\phi^i)\partial_\mu\phi^p\partial_\nu\phi^q \,, \tag{2}$$

where the scalar manifold metric $h_{ij}(\phi)$, with $i, j = 1, 2$, can be put in the conformally flat form

$$h_{ij}(\phi) = \frac{4}{[1-(\phi^1)^2-(\phi^2)^2]^2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} , \tag{3}$$

and where the scalar fields are restricted to the noncompact unit disc $0 \leqslant (\phi^1)^2 + (\phi^2)^2 < 1$. In order to solve the equations, we assume that two of the spatial dimensions form a surface topologically equivalent to H_2. We can then set the background values of the scalar fields equal to differentiable invertible functions of the coordinates in question. Without loss of generality, we can take $\phi^i(x^\mu) = x^i$. This equality [6,7] has the effect in the present case of restricting x^1 and x^2 to the noncompact space with $0 \leqslant (x^1)^2 + (x^2)^2 < 1$, and reduces eq. (1) for the background metric $\bar{g}$ to the equations

$$R_{ij}(\bar{g}) = \tfrac{1}{2} h_{ij}(x^k) \,, \quad R_{\alpha\beta}(\bar{g}) = 0 \,, \tag{4,5}$$

where i, j are curved indices for the extra dimensions and α, β are curved indices for the remaining dimensions. We have a product space where the factor space for the remaining dimensions has no cosmological constant, as we see from eq. (5).

A short computation shows that if eq. (4) is satisfied then eq. (2) contains no further information [8]. In eq. (4) we are asked to find a metric $\bar{g}_{ij}$ such that its Ricci curvature equals the scalar manifold metric. It is easy to see that equating the two metrics $g = h$, as in refs. [6,7], will not do in the present case because h_{ij} is a metric of constant negative curvature and therefore the left-hand side of eq. (4) would be negative definite in contradistinction to the right-hand side of the equation which is manifestly positive.

In order to find the solution of eq. (4) one uses the fact that two-dimensional spaces are conformally flat. One can set $\bar{g}_{ij} = \Omega^2\eta_{ij}$, where Ω^2 is an unknown conformal factor and $\eta_{ij} = \mathrm{diag}(1, 1)$. (Actually, this is the first use we are making of the restriction to two dimensions; for larger numbers of dimensions there may be other kinds of solutions of eq. (4).) Eq. (4) then gives a differential equation for Ω^2 that can be solved to find

$$\mathrm{d}s^2 = \bar{g}_{ij}\,\mathrm{d}x^i\mathrm{d}x^j = 4a^2(1-x^2-y^2)(\mathrm{d}x^2+\mathrm{d}y^2) \,, \tag{6}$$

where the size parameter a^2 is not determined by the equations of motion. The metric given above describes the shape of the curled-up two-dimensional surface. The value of the self-coupling strength in the nonlinear sigma model has determined the power $(1-x^2-y^2)$ in eq. (6).

The surface, although topologically noncompact (see for example ref. [9]), has finite area since the integral $\int\sqrt{g}\,\mathrm{d}^2x$ for $|\boldsymbol{x}| < 1$ exists. The scalar curvature R behaves as $(1-|\boldsymbol{x}|^2)^{-3}$ so that the space becomes more and more curved as $|\boldsymbol{x}| \to 1$. Given that $|\boldsymbol{x}| = 1$ does not belong to the space, the curvature is never infinite, although it can be arbitrarily large.

In order to visualize the surface, one can try to embed it isometrically in three-dimensional euclidean space. The part of the surface corresponding to $0 \leqslant |\boldsymbol{x}|^2 < 2/3$ can be embedded in ordinary euclidean space with signature $(+++)$ as shown in fig. 1. The bottom of the surface is at $\boldsymbol{x} = 0$ with finite curvature. The surface opens up until point B ($|\boldsymbol{x}|^2 = 1/2$) is reached. Then it starts to close and at point C ($|\boldsymbol{x}|^2 = 2/3$) its slope is zero. Beyond point C this embedding breaks down. The rest of the surface can be embedded (at the price of losing some visualization) in three-dimensional euclidean space with signature $(++-)$[‡1], as shown in dotted lines in fig. 1. There is a cusp at point D corresponding to $|\boldsymbol{x}| = 1$, which, as mentioned before, does not belong to the surface. The tear drop has a compact U(1) isometry group, i.e. rotational symmetry.

Despite the unusual shape of the two-dimensional surface and the fact that it is not an Einstein space, we can verify that the violation of supersymmetry is not total. In the case of the chiral $N = 2, D = 10$

[‡1] There may exist other instructive embeddings. For example, the surface of constant negative curvature H_2 can be thought of as a hyperboloid entirely embedded in $(++-)$, but it is also possible to embed a part of H_2 isometrically in $(+++)$ and the rest in $(++-)$. We wish to thank O. Alvarez for useful discussions on embeddings.

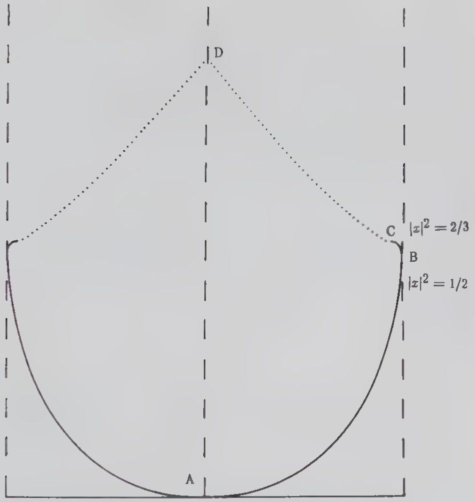


Fig. 1. Embedding of the two-dimensional surface.

theory the scalar-gravitational background admits a global supersymmetry. In this theory, the gravitino and the spinor remain invariant under supersymmetry [3] provided there is a Killing spinor ϵ^I satisfying the conditions

$$\delta\psi_\mu \sim (\partial_\mu + \tfrac{1}{4}\omega_{\mu rs}\Gamma^{rs} - \tfrac{1}{2}iQ_\mu)\epsilon^I = 0 \,, \tag{7}$$

$$\delta\chi \sim \Gamma^\mu P_\mu \epsilon^{*I} = 0 \,, \tag{8}$$

where, according to the conventions of ref. [3], $\omega_{\mu rs}$ is the spin connection, $Q_\mu = \mathrm{Im}\,(z\partial_\mu z^*)/(1-zz^*)$, $P_\mu = \partial_\mu z/(1-zz^*)$, and $z = \phi^1 + i\phi^2$. Eq. (8) requires that

$$\Gamma^{12}\epsilon^I = i\epsilon^I \,, \tag{9}$$

which halves the number of supersymmetries from 16 Weyl-supercharges to 8 Weyl-supercharges. If condition (9) is satisfied, the two connections ω and Q in eq. (7) cancel against each other and one is left with $\partial_\mu \epsilon^I = 0$, so that ϵ^I is coordinate-independent. Surviving global supersymmetry strongly suggests stability of the background solution [10,11]. Although we have not carried out a complete analysis of the problem of small fluctuations, we expect the conclusions to be analogous to those reached in the context of anti-de Sitter backgrounds in gauged supergravity. In that case boundary conditions are put on the fluctuations at spatial infinity [10]. These boundary conditions prevent energy from escaping through spatial infinity, produce quantization of the spectrum, give rise to a well-defined Cauchy problem, and are compatible with supersymmetry [10,12]. In the present problem, boundary cndoitions should be imposed for $|\boldsymbol{x}| \to 1$ in order to prevent energy from escaping through the hole at the cusp of the surface. A partial discussion of the fluctuation problem is given in ref. [8].

Our solution implies that the $N = 2, D = 10$ theory reduces to eight dimensions with $N = 1$ supersymmetry (corresponding to $N = 4$ supersymmetry in four dimensions) and thus the zero mass sector should correspond to an $N = 1, D = 8$ supergravity coupled perhaps to some super-Yang–Mills matter multiplets. $N = 1$ supergravity in eight dimensions contains 48 + 48 modes and describes a graviton, two vectors, a two-form, a scalar, a Weyl gravitino, and a Weyl spinor. The $D = 8$ super-Yang–Mills multiplet contains 8 + 8 modes and describes a vector, a Weyl spinor, and two scalars. A detailed analysis on the tear drop of the zero modes of the various differential operators corresponding to the various fields of chiral $N = 2, D = 10$ supergravity is necessary in order to determine whether or not there are massless super-Yang–Mills multiplets. At any rate further reduction to four dimensions is possible. If one insists on not inducing a cosmological constant at the classical level [‡2], one may consider compactification over four tori, yielding $N = 4$ supergravity coupled to at least four super-Yang–Mills multiplets as the zero-mass sector. Another possibility is compactification of four dimensions over K3 [15,16], yielding – the relevant counting was worked out in ref. [15] – $N = 2$ supergravity coupled to one super-Yang–Mills multiplet and at least 20 hypermultiplets. Given that these theories have no massless vector gauge fields, one would hope that such fields arise dynamically, perhaps in the way Cremmer and Julia [17] suggested for $N = 8$ supergravity or in the way recently suggested by Mandelstam [18].

Dimensional reduction induced by H_2 is possible for several other higher-dimensional supergravities.

[‡2] This may not be necessary if spacetime foam ideas are correct [13]. One could then contemplate compactification induced by vector fields or antisymmetric tensors [14].

For example, $N = 2, D = 9$ supergravity has 3 scalars living on $R \times H_2$ [19], so that reduction to $D = 7$ is possible. Similarly, the $N = 2, D = 8$ theory has 7 scalars living on $(SL(3,R)/SO(3)) \times H_2$ [19], so that reduction to $D = 6$ can be accomplished. Also, $N = 1$, $D = 8$ supergravity coupled to two super-Yang–Mills multiplets [‡3] (which can be obtained by reduction of $N = 1, D = 10$ supergravity over two tori) has 5 scalars living on $R \times H_2 \times H_2$, so that reduction down to $D = 4$ is possible with the aid of two tear drops [‡4].

Novel as they are, noncompact spaces such as the tear drop discussed here (or higher-dimensional analogues that may exist) are interesting alternatives to the usual compact manifolds used in spontaneous compactification. In fact, noncompact internal spaces have been advocated very recently by Wetterich [20] as a way to by-pass the "no-go theorems" [21] on chiral fermions for compact manifolds. Of course the theorems apply only to internal spaces with continuous isometries and so compactification over K3, for example, can preserve left–right asymmetry [16]. In the case of the tear drop, the surface has a U(1) isometry group and nevertheless the theorems do not apply. If there are gauge vector fields with suitable couplings, a kind of left–right asymmetry could be preserved by our reduction from ten to eight dimensions. It is not so clear, however, how to exploit this situation, since the Kaluza–Klein vector field associated with the isometry, which could have chiral couplings to massless fermions, acquires a mass at the classical level [8], as described for the case of compact sigma models in refs. [6,7].

In the conventional approach to dimensional reduction, one seeks surviving four-dimensional massless fermions with chiral couplings to gauge vector fields arising from a higher-dimensional metric tensor. In order to achieve that, it is necessary that mass terms be forbidden by generalized Lorentz invariance for some of the fermions in the higher-dimensional theory [22]. This condition, that some fermions in the higher-dimensional theory have their masses forbidden without reliance on supersymmetry, may be important even apart from the conventional approach. Chiral $N = 2, D = 10$ supergravity obeys the condition, and the tear drop reduction to eight dimensions does not wash it away, at least for the gravitino. Reduction over two tori would, however, violate the condition.

From many points of view, dimensional reduction over noncompact manifolds opens up exciting new possibilities.

References

[1] E. Cremmer, J. Scherk and S. Ferrara, Phys. Lett. 68B (1977) 234.

[2] M.B. Green and J.H. Schwarz, Phys. Lett. 122B (1983) 143;
J.H. Schwarz and P.C. West, Phys. Lett. 126B (1983) 301.

[3] J.H. Schwarz, Nucl. Phys. B226 (1983) 269.

[4] P. Howe and P.C. West, Nucl. Phys. B238 (1984) 181.

[5] J.H. Schwarz, Phys. Rep. 89 (1982) 223;
M.B. Green, Surv. High Energy Phys. 3 (1983) 127;
M.B. Green and J.H. Schwarz, Caltech preprint CALT-68-1090.

[6] C. Omero and R. Percacci, Nucl. Phys. B165 (1980) 351.

[7] M. Gell-Mann and B. Zwiebach, Phys. Lett. 141B (1984) 333.

[8] M. Gell-Mann and B. Zwiebach, to be published.

[9] R. Gilmore, Lie groups, Lie algebras and some of their applications (Wiley Interscience, New York, 1974).

[10] P. Breitenlohner and D.Z. Freedman, Phys. Lett. 115B (1982) 197; Ann. Phys. (NY) 144 (1982) 249.

[11] G.W. Gibbons, C.M. Hull and N.P. Warner, Nucl. Phys. B218 (1983) 173;
C.M. Hull, Commun. Math. Phys. 90 (1983) 545.

[12] S.W. Hawking, Phys. Lett. 126B (1983) 175.

[13] S.W. Hawking, Nucl. Phys. B144 (1978) 349; Talk Shelter Island Conf. (June 1983).

[14] P.G.O. Freund and M.A. Rubin, Phys. Lett. B97 (1980) 233.

[15] M.J. Duff, B.E.W. Nilsson and C.N. Pope, Phys. Lett. 129B (1983) 39.

[16] P.K. Townsend, Phys. Lett. 139B (1984) 283.

[17] E. Cremmer and B. Julia, Nucl. Phys. B159 (1979) 141.

[18] S. Mandelstam, Talk Los Alamos Conf. (December 1983).

[19] B. Julia, Superspace and supergravity, eds. S.W. Hawking and M. Roček (Cambridge U.P., London, 1980).

[20] C. Wetterich, University of Bern preprints BUTP-84/5, BUTP-84/6 (1984).

[21] E. Witten, Fermion quantum numbers in Kaluza–Klein theory, Princeton preprint (October 1983) to be published in Nucl. Phys. B.

[22] C. Wetterich, Nucl. Phys. B211 (1984) 177.

[‡3] This example was suggested to us by J.H. Schwarz.

[‡4] It would be amusing if this last theory turned out to be the zero-mass sector arising from the curling up into a tear drop of two dimensions in chiral $N = 2, D = 10$ supergravity; three tear drops would then reduce ten dimensions to four. (Here we are including the possibility of modified tear drops with different powers of $1 - x^2$ in the metric, in case we have different self-coupling strengths in the H_2 sigma models.)

CHAPTER
9

Monopoles

Kaluza-Klein Monopole

Rafael D. Sorkin

Institute for Advanced Study, Princeton, New Jersey 08540, and Center for Theoretical Physics,[a]
University of Maryland, College Park, Maryland 20742

(Received 18 April 1983)

By adding the trivial term $-dt^2$ one can convert the four-dimensional (positive-definite) Newman-Unti-Tamburino line element into a static solution of the five-dimensional vacuum Einstein equations. Interpreted *à la* Kaluza and Klein, the solution describes a monopole with a charge-to-mass ratio of $\sqrt{2}$ in rationalized Planck units. Although it is a perfectly regular five-geometry it appears singular from a four-dimensional perspective.

PACS numbers: 12.10.-g, 04.20.Jb, 04.60.+n, 14.80.Hv

Present-day physics exhibits two independent kinds of symmetries: the geometrical symmetries expressing the interchangeability of the points of space-time, and the internal symmetries expressing an interchangeability among certain of the material structures which can occur at a single space-time point. The aim of the so-called Kaluza-Klein theories is to unify these types of symmetry by hypothesizing that the apparently internal symmetries are in reality also geometrical. One posits the existence of extra (but very compact) dimensions of space-time and interprets gauge transformations as generalized rotations taking place within these extra dimensions.

In the prototype Kaluza-Klein theory[1] the extra dimension was no more than a mathematical artifice since the internal metric (that of a circle) was fixed *a priori*, and not to be varied in the action principle. When this restriction is dropped excitations of the internal geometry become possible, and the extra dimensions can have observable effects.

Nevertheless a cynic might still choose to regard the use of such dimensions as a trick for producing interesting four-Lagrangians. Short of direct exploration of the "internal" geometry, the best antidote to such cynicism might be the discovery of a particle whose hypothesized structure *does not admit* of interpretation in terms of four-dimensional fields. In fact such structures exist, and we will see below that one which occurs in the purely electromagnetic theory can be interpreted as a magnetic monopole carrying the Dirac charge.

The Lagrangian.—For the Abelian [U(1)] Kaluza-Klein theory, space-time is a five-manifold ${}^5M$ with metric g_{AB} of signature $-++++$. To descend to four dimensions one introduces a vector field K^A whose orbits are circles ($\exp 2\pi K = 1$) and assumes that, at least to a first approximation, K is a Killing vector for g_{AB}. The five-metric then decomposes as

$$g_{AB} = \gamma_{AB} + A_A A_B , \tag{1}$$

where γ_{AB} is a degenerate metric of signature $0-+++$ such that $\gamma_{AB}K^B = 0$. The one-form A is well defined wherever $K \neq 0$ and related to K^A by $A_A = K_A/\lambda = g_{AB}K^B/\lambda$, where $\lambda = (g_{AB}K^AK^B)^{1/2}$ is the radius of the "internal circle."

The effective four-manifold ${}^4M$ is now the *quotient* of ${}^5M$ with respect to the action of K: Each circular orbit of K becomes a single point in ${}^4M$. Because of the assumption that K is a symmetry, nothing really depends on the fifth dimension; and g_{AB} is equivalent to a set of three distinct fields defined on ${}^4M$, a scalar, a tensor, and a vector, deriving respectively from λ, γ_{AB}, and A_A. At spatial infinity λ will approach a constant value λ_∞, and we take the four-scalar field to be

$$\omega = -\tfrac{1}{2}\ln(\lambda/\lambda_\infty) . \tag{2}$$

To define the four-metric $\tilde{g}_{\mu\nu}$ (using Greek indices for ${}^4M$) notice that there is a unique metric $\gamma_{\mu\nu}$ whose pullback to ${}^5M$ is γ_{AB} and put

$$\tilde{g}_{\mu\nu} = (\lambda/\lambda_\infty)\gamma_{\mu\nu} = e^{-2\omega}\,\gamma_{\mu\nu} . \tag{3}$$

In a coordinate system with x^5 parametrizing the internal dimension and $K^A = (0, 0, 0, 0, 1)$ the components of g_{AB} are independent of x^5 and our definitions so far read $\lambda = (g_{55})^{1/2}$, $\tilde{g}_{\mu\nu} = (\lambda/\lambda_\infty)[g_{\mu\nu} - g_{\mu5}g_{\nu5}/g_{55}]$. Finally we will need the vector potential $\tilde{A}_\mu \equiv A_\mu/\lambda$. Unlike ω and $\tilde{g}_{\mu\nu}$, $\tilde{A}_\mu$ is gauge dependent—it changes under coordinate transformations $x^5 \to x^5 + \theta(x^\mu)$—but of course

$$\tilde{F}_{\mu\nu} \equiv (\mathrm{curl}\tilde{A})_{\mu\nu} \equiv \partial_\mu \tilde{A}_\nu - \partial_\nu \tilde{A}_\mu \tag{4}$$

is not, and can be defined invariantly as the two-form on ${}^4M$ whose pullback to ${}^5M$ is $F \equiv \mathrm{curl}(A/\lambda)$.

With these definitions[2] the five-Lagrangian $(2k)^{-1}({}^5R)(-{}^5g)^{1/2}$ leads to a four-action

$$S = S_\infty + \kappa^{-1} \int (\tfrac{1}{2}\tilde{R} - 3\,|\tilde{\nabla}\omega|^2 - \tilde{\nabla}^2\omega - e^{-6\omega}\lambda_\infty{}^2/8\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu})d\tilde{V}\,, \tag{5}$$

where $\tilde{R}$ is the scalar curvature of $\tilde{g}_{\mu\nu}$ (with respect to which all contractions are done), and $\kappa = k/2\pi\lambda_\infty$ plays the role of the four-dimensional gravitational constant $8\pi G$. The integral S_∞ at spatial infinity would be needed in deducing the total energy from first principles, but we may neglect it for what follows.

A monopole solution.—Since the five-Lagrangian is just the scalar curvature the five-dimensional field equations are just ${}^5R_{AB} = 0$. Consequently any signature-(++++) solution g_{AB} of four-gravity can be extended to a static Kaluza-Klein solution, g_{AB}, simply by setting $g_{00} = -1$, $g_{a0} = 0$ (where $a, b = 1, 2, 3, 5$). For example the flat metric derived from $g_{ab} = \delta_{ab}$ by periodically identifying $x^5 = x^5 + 2\pi$ gives rise in this way to the Kaluza-Klein equivalent of Minkowski space-time.

Applying this procedure to the so-called Taub-Newman-Unti-Tamburino (NUT) instanton yields a five-metric which can be written conveniently[3] in terms of Cartesian coordinates t, x, y, z, w for R^5. In terms of $q^2 = x^2 + y^2 + z^2 + w^2$, $\rho = 2\ln q$, and $\sigma_z = 2q^{-2}(z\,dw - w\,dz + x\,dy - y\,dx)$ (with cyclic permutations for σ_x and s_y), the full Kaluza-Klein metric is

$$ds^2 = -dt^2 + U^{-1}d\rho^2 + 4m^2U\sigma_z{}^2 + (\rho^2 - m^2)(\sigma_x{}^2 + \sigma_y{}^2)\,, \tag{6}$$

where $U = (\rho - m)/(\rho + m)$. In (6) there is an apparent singularity at $\rho = m$. However, a simple radial coordinate transformation and redefinition of ρ brings ds^2 to a form valid for all t, x, y, z, w, and in particular shows that (for any $m > 0$) the topology of the solution is simply R^5. The metric g_{AB} thereby assumes the form[4]

$$-\partial t^2 + \left(1 + \frac{m}{\rho}\right)^2 \partial r^2 + \left(\frac{2mr}{\rho + m}\right)^2 \sigma_z{}^2 + r^2(\sigma_x{}^2 + \sigma_y{}^2)\,, \tag{7}$$

where $r \equiv x^2 + y^2 + z^2 + w^2$ and

$$\rho^2 \equiv r^2 + m^2\,. \tag{8}$$

An equivalent expression in terms of Euler angles for the three-sphere can be obtained by substituting[3]

$$\sigma_z = -\partial\psi - \cos\theta\,\partial\varphi\,, \tag{9}$$

$$\sigma_x{}^2 + \sigma_y{}^2 = \partial\theta^2 + \sin^2\theta\,\partial\varphi^2 \equiv \partial\Omega^2\,, \tag{10}$$

with $0 \leqslant \theta \leqslant \pi$ and with φ and ψ having periods 2π and 4π, respectively.[5] That (7) is regular at $r = 0$ may not be immediately recognizable, but direct computation reveals that near $r = 0$ it reduces to $-\partial t^2 + 4(\partial w^2 + \partial x^2 + \partial y^2 + \partial z^2)$, whose smoothness is manifest.

Equation (7) defines the five-dimensional soliton solution. It is manifestly static, already in the form (1) with $A = 2mr/(\rho + m)\sigma_z$, and in $t, r, \theta, \varphi, \psi$ coordinates clearly admits $K = 2\,\partial/\partial\psi$ as Killing vector with period 2π (i.e., $\psi/2$ can be used to parametrize the internal circles[6]). It follows that

$$\lambda = 4mr/(\rho + m)\,. \tag{11}$$

For $r \gg m$ the metric approaches $-\partial t^2 + \partial r^2 + r^2\partial\Omega^2 + 4m^2\sigma_z{}^2$, which locally looks like a "Kaluza-Klein Minkowski space" with internal radius $8\pi m$; in particular $\lambda_\infty = 4m$.

Proceeding from these formulas it is easy to pass to the four-description of the metric. The quotient manifold ${}^4M$ is naturally coordinatized by t, r, θ, φ and is thus topologically R^4. The true space-time ${}^5M$ is everywhere a fiber bundle over ${}^4M$ *except at the origin* $r = 0$, where $\lambda = 0$ and the U(1) fiber shrinks to a point. From (10) $\gamma_{\mu\nu}$ is $-\partial t^2 + (1 + m/\rho)^2\,\partial r^2 + r^2\,\partial\Omega^2$, whence the quotient metric $\tilde{g}_{\mu\nu}$ of (3) is [with the help of (8)]

$$-[(\rho - m)/r]\,\partial t^2 + (r/\rho)(1 + m/\rho)\partial r^2 + (\rho - m)r\,\partial\Omega^2\,. \tag{12}$$

From (2) and (11) we see that

$$\omega = \tfrac{1}{2}\ln[(\rho + m)/r]\,. \tag{13}$$

Finally $\mathrm{curl}(A/\lambda) = \tfrac{1}{2}\mathrm{curl}\,\sigma_z = \tfrac{1}{2}(\sin\theta\;\partial\theta \wedge \partial\varphi)$ which translates immediately to ${}^4M$ to give for the electromagnetic field tensor

$$\tilde{F} = \tfrac{1}{2}\sin\theta\,\partial\theta \wedge \partial\varphi\,. \tag{14}$$

Observe that $\tilde{g}_{\mu\nu}$, ω, and $\tilde{F}$ all become singular at $r = 0$, as could have been anticipated from the breakdown there of the local product structure of ${}^5M$.

Mass and charges of the solution.—To justify calling this solution a monopole let us compute its electromagnetic charges. The total magnetic flux through any sphere about the origin is, from (14),

$$\oint \tilde{B} \cdot ds = \oint \tfrac{1}{2} \tilde{F}_{\mu\nu}\, d\Sigma^{\mu\nu} = \int \tfrac{1}{2} \sin\theta\, d\theta\, d\varphi = 2\pi ,$$

as could have been inferred directly from the fact that the NUT metric has one twist[7] when regarded as a U(1) bundle over ${}^4M$. However, this flux is not immediately the magnetic charge because the coefficient of $F_{\mu\nu}F^{\mu\nu}$ in (5) differs from the customary $-\frac{1}{4}$, the correctly normalized magnetic charge[8] being rather $Q_B = 2\pi\lambda_\infty \times (2\kappa)^{-1/2}$. In units such that $k = 2\pi\lambda_\infty = 1$ (note, $\hbar \neq 1$!), which I designate "KK units," this becomes $Q_B = 1/\sqrt{2}$.

As for the total *electric* flux, $\oint \tilde{E} \cdot dS$, it vanishes because the absence of ∂t in (14) means that E itself vanishes. Notice incidentally that the Dirac condition $eQ_B/h =$ integer (where $h = 2\pi\hbar$) is verified with $n = 1$ since the unit of electric charge is $e = (2\kappa)^{1/2}\hbar/\lambda_\infty$.

The monopole's scalar charge is somewhat easier to compute. According to (13) and (8) $\omega \approx m/2r$ at large radii and is therefore the field of a source of strength $2\pi m = (\pi/2)\lambda_\infty$. Again, however, ω should be rescaled by $(6/\kappa)^{1/2}$ to bring its contribution to (5) to the usual form. The resulting scalar charge is $Q_S = (3/2\kappa)^{1/2}\pi\lambda_\infty$ or $(\frac{3}{8})^{1/2}$ in KK units.

Finally, what about the mass? With the definitions of this paper we can deduce the total energy M unambiguously[9] from the behavior of $\tilde{g}_{00}$ as $r \to \infty$. From (12) $-\tilde{g}_{00} \sim 1 - m/r$, whence $M = 4\pi m/\lambda_\infty = \pi\lambda_\infty/\kappa$ or $\frac{1}{2}$ in KK units. To summarize, then, the mass and charges in KK units are[10]

$$M = \tfrac{1}{2}, \quad Q_B = (\tfrac{1}{2})^{1/2}, \quad Q_S = (\tfrac{3}{8})^{1/2} .$$

Incidentally the value of Q_B implies that bound states uniting the monopole with charge e excitations of g_{AB} would be fermions.[11]

Discussion.—In no four-dimensional theory with a scaling symmetry can there be a stationary solution of nonzero energy.[12] How then do the Kaluza-Klein equations admit a solution (7) which is not only stationary, but everywhere regular and without event horizon? Two features seem essential. The inescapable five-dimensionality of the metric prevents any four-theorems from being applied directly; and the fixed value of the internal radius λ_∞ in effect introduces a fundamental length which then prevents the five-dimensional analog theorem from applying. This seems quite similar to the way in which a boundary condition of the Higgs field allows monopole solutions to persist in the Prasad-Sommerfield limit, even though the scale-invariance-breaking potential itself goes to zero. Here, however, there is no Higgs field and no symmetry breaking.

In the same way as we used the NUT metric, each of the known ALF instantons could be used to construct a static Kaluza-Klein solution. For example the multi-Taub-NUT solutions become multimonopole solutions and the Schwarzschild instanton of "mass" M yields a neutral solution of total energy $M/2$ which seems to describe a delicately balanced "bubble" of the type discussed by Witten.[13] By a slight generalization of the present procedures (to allow g_{0A} to vary) the Israel-Wilson solutions may also turn into five-dimensional vacuum solutions.

Of course all these solutions are physically unrealistic insofar as U(1) is not the gauge group realized in nature.[14] But in order that there be analogs it may suffice that the "correct" G contain U(1) as a subgroup, which is virtually certain.

In a way it is surprising that so many seemingly distinct particlelike structures can exist in what seem to be fundamental theories.[15] One might wish that at least some of these possibilities could be, if not excluded, then at least reduced to some of the others. In the context of the present theory, for example, there is no reason why the charge carried by the monopole should be labeled "magnetic." It might equally well be deemed "electric" and $\tilde{A}_\mu$ treated as a potential for the dual field ${}^*F_{\mu\nu}$. With this reidentification— or even without it—it becomes important to understand how the effective dual potential needed to express the mutual interaction of such monopoles can arise. Once this is done,[16] it may appear that some or all of what today seem to be elementary structures (including space-time itself) arise in a similar way from "sub-Planck-scale" quantum processes.

For interesting comments I would like to thank those who attended the Institute for Advanced Study-Princeton University relativity seminar at which I first presented this idea, including especially Malcolm Perry. I would also like to thank Paul Green and other University of Maryland colleagues for suggestions.

[(a)] Present address.

[1] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Math. Phys. 1921, 966; O. Klein, Z. Phys. 37, 895 (1926).

[2] And with the following conventions: $c \equiv 1$, $[\nabla_M, \nabla_N]V^A = R^A{}_{BMN}V^B$, $R_{MN} = R^A{}_{MAN}$.

[3] The Taub-NUT metric is described by A. Taub and C. W. Misner, Zh. Eksp. Teor. Fiz. 55, 233 (1968) [Sov. Phys. JETP 28, 122 (1969)]. The present metric corresponds to the choice $l = im$ (with "ρ" denoting their coordinate t).

[4] Here "∂t" denotes the one-form $\partial_A t$; "∂t^2" stands for $\partial t \otimes \partial t$, i.e., $\partial_A t\, \partial_B t$; and σ_x, σ_y, and σ_z are to be understood as one-forms defined in R^4.

[5] To remove the familiar polar coordinate singularities one makes the unfamiliar extra identifications at $\theta = \pm\pi$, $(\theta, \varphi, \psi) = (\theta, \varphi + \Delta, \psi \mp \Delta)$, Δ arbitrary.

[6] Except near $\theta = \pi$ where $\psi - \varphi$ can be used.

[7] S. Ramaswamy and A. Sen, J. Math. Phys. 22, 2621 (1981).

[8] In so-called Heaviside units for which the force between two charges is $Q_1 Q_2/4\pi r^2$.

[9] There can be only one conserved quantity whose "matter" contributions (from the stress-energy tensors for ω and $F_{\mu\nu}$) are correctly normalized in regions of small gravitational field. It was for this convenience in identifying M that I chose to use (3) and not just γ as the four-metric.

[10] Thus the present solution has an electromagnetic charge greater than that of any Reissner-Nordstrom metric of the same mass.

[11] See A. O. Barut, Phys. Rev. D 10, 2712 (1974); R. D. Sorkin, Phys. Rev. D 27, 1787 (1983).

[12] This follows from a theorem of B. F. Schutz and R. D. Sorkin, Ann. Phys. (N.Y.) 107, 1 (1977), with the assumption that the theory is defined by an "unconstrained" action principle.

[13] E. Witten, Nucl. Phys. B195, 481 (1982). Notice though that with the metric $\tilde{g}_{\mu\nu}$ the bubble becomes a point singularity and there is actually no "hole" in the apparent space-time.

[14] They are probably also quantum mechanically unstable via the mechanism of Ref. 13. However, one should at least decide which ones are classically stable, say with the methods of G. W. Gibbons and C. M. Hull, Phys. Lett. 109B, 190 (1982).

[15] For some examples see E. Lubkin, Ann. Phys. (N.Y.) 23, 233 (1963); R. D. Sorkin, J. Phys. A 10, 717 (1977), and 11, 795 (1978); D. Finkelstein, J. Math. Phys. 7, 1218 (1966).

[16] B. Hu, in *Proceedings of the Guangzhou Conference on Particle Physics* (Science Press, Beijing, 1980).

MAGNETIC MONOPOLES IN KALUZA–KLEIN THEORIES*

David J. GROSS and Malcolm J. PERRY

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544, USA

Received 12 April 1983

We demonstrate that the five-dimensional Kaluza–Klein theory of unified gravity and electromagnetism admits soliton solutions. These are regular, static and stable solutions of the field equations which correspond, upon quantization, to particles. The solitons include magnetic monopoles, which obey the Dirac quantization condition, as well as magnetic dipoles which are topologically stable. The inertial mass of the solitons is typically of order m_P/e, where m_P is the Planck mass and e the electric charge. These solitons have bizarre gravitational interactions; in fact they exert no newtonian force on slowly moving test particles, thus they have zero gravitational mass. We explain how the inequality of the gravitational and inertial masses is due to the violation of Birkhoff's theorem in Kaluza–Klein theories and is consistent with the principle of equivalence.

1. Introduction

One of the oldest and most enduring ideas regarding the unification of gravitation and gauge theory is Kaluza's five-dimensional unified theory [1]. Kaluza's hypothesis was that the world has four spatial dimensions, but that one dimension is curled up to form a circle so small as to be unobservable. He showed that ordinary general relativity in five dimensions, assuming such a cylindrical ground state, contained a local U(1) gauge symmetry arising from the isometry of the hidden fifth dimension. The extra components of the metric tensor constituted the gauge fields of this symmetry and could be identified with the electromagnetic vector potential.

This theory was investigated by many authors including Klein, who greatly clarified its structure, and others [2]. Its extension to non-abelian gauge theories was anticipated by Klein, and developed by de Witt, Kerner, Trautman, Cho and Freund [3]. The basic structure of the generalized Kaluza–Klein theories is the following. One considers a world consisting of d dimensions (1 timelike and $d-1 \geqslant 3$ spacelike). The lagrangian is simply the Einstein–Hilbert action in d dimensions,

$$L = \frac{1}{16\pi G_K} \int d^d x \sqrt{-g}\, R\,,$$

although additional matter fields or supergravitational actions can be considered.

---

* Supported in part by NSF grant PHY80-19754.

One then assumes that the ground state or vacuum is the product of four-dimensional Minkowski space with an internal, compact space, $I^{(d-4)}$. This ground state is symmetric under four-dimensional Poincaré transformations and the isometry group, G, of $I^{(d-4)}$. Associated with each generator of the internal isometry group there exists a local gauge symmetry and a corresponding massless gauge boson. The resulting theory, which describes physics at energies low compared to the inverse size of the internal dimensions, is ordinary gravity coupled to a Yang–Mills theory with gauge group G.

Kaluza–Klein theories have been the subject of revived interest in recent years, for a variety of reasons. First it is now clear that all non-gravitational interactions are dictated by gauge symmetries, and that the natural scale of their unification is reasonably close to the Planck mass where gravitation can no longer be neglected. Also we have become quite accustomed to the phenomenon of spontaneous symmetry breaking, and can well imagine, as Kaluza assumed, that the vacuum is not maximally symmetric due to "spontaneous compactification" [4]. Finally the emergence of supersymmetric gravitation has raised the possibility of providing a geometric theory of both gauge interactions and fermionic matter.

To be sure there are many substantial difficulties in realizing this goal. We do not yet possess a quantum mechanical theory of gravity which can adequately describe gravitational interactions at short distances. We do not have a compelling reason to expect spontaneous compactification of the internal dimensions, particularly in view of our ignorance concerning the cosmological constant. Finally there exists no realistic Kaluza–Klein theory, which is consistent with our knowledge of low-energy physics [5]. Nonetheless, we sympathize with the point of view expressed by Einstein and Pauli [6] that one "cannot help feeling that there is some truth in Kaluza's five-dimensional theory." Even in the absence of solutions to the above problems it is important to explore the structure of Kaluza–Klein theories. Most important are attempts to derive general consequences of these theories that might be susceptible to experimental test.

In this paper we shall explore the properties of solitons in the original, five-dimensional, Kaluza–Klein theory. Solitons, i.e. non-singular, topologically stable, solutions of the field equations which describe localized lumps of matter, have long been sought in general relativity, notably by Einstein*. No such solitons exist in ordinary four-dimensional relativity. On the other hand it is easy, as we shall see, to find solitons in the Kaluza–Klein theory. The simplest soliton is in fact a magnetic monopole, which is hardly surprising in a unified theory of electromagnetism and gravity. However there exist as well topologically stable magnetic-dipole solitons. We shall explicitly construct these solutions, discuss their topological properties, calculate their masses, show how magnetic charge quantization emerges, and investigate their (rather bizarre) interactions. We discuss, in this paper, only the simplest five-dimensional Kaluza–Klein theory, although many of our considerations should be generic to other Kaluza–Klein theories.

---

* Einstein attempted to find solitons in the Kaluza–Klein theory, and in fact argued that [6, 7] there can exist no such objects. The solutions exhibited in this paper contradict these arguments, which mistake coordinate singularities for real singularities.

In sect. 2 we review the structure of the Kaluza–Klein theory. In particular we discuss the form of the effective, four-dimensional, theory that describes low-energy physics. This is a Brans–Dicke theory of coupled gravity and electromagnetism with a massless scalar coupled in a specific way to matter.

Sect. 3 is devoted to the construction of the primary monopole solution. We analyze its topological character, evaluate its inertial mass and show how the Dirac quantization condition emerges. We also construct multi-monopole solutions, corresponding to a collection of noninteracting monopoles.

The monopoles, in addition to magnetic charge, are characterized by the topology of their spatial sections. Both monopoles and anti-monopoles carry one unit of Euler character (the, fixed time, spatial manifolds have a "handle") and thus cannot classically annihilate. Therefore one can construct stationary magnetic dipole solutions. This is done in sect. 4.

In sect. 5 we discuss the strange gravitational properties of the solitons. We find that they exert no force on a neutral test particle at rest: their gravitational mass is zero. We explain how this is consistent with the principle of equivalence and arises from the interaction of matter with the Brans–Dicke scalar field. Finally we summarize our findings in sect. 6, draw some general conclusions and raise some unanswered questions

2. The Kaluza–Klein theory

The original Kaluza–Klein theory is simply five-dimensional general relativity, under the assumption that one spatial dimension is curled up in a circle, whose radius is so small as to be unobservable. The action is

$$S = -\frac{1}{16\pi G_{\mathrm{K}}} \int \mathrm{d}^5 x \sqrt{-g_5}\, R_5 \,, \tag{1}$$

where $R_5(g_{AB})$ is the five-dimensional curvature scalar of the metric g_{AB}, and G_{K} is the five-dimensional coupling constant. Our conventions are: upper case Latin letters A, B, C denote five-dimensional indices 0, 1, 2, 3, 5; lower case Greek indices run over four dimensions 0, 1, 2, 3 whereas lower case Latin indices run over four-dimensional spatial values 1, 2, 3. The signature of g_{AB} is $-++++$, the Riemann tensor is

$$R^K{}_{LMN} = \partial_M \Gamma^K{}_{LN} - \partial_N \Gamma^K{}_{LM} + \Gamma^K{}_{JM} \Gamma^J{}_{LN} - \Gamma^K{}_{JN} \Gamma^J{}_{LM} \,,$$

$$R_{LM} = R^K{}_{LKM} \,, \qquad R_5 = R^L{}_L \,,$$

and except where indicated $\hbar = c = 1$.

In the absence of other fields the equations of motion are of course $R_{AB}=0$. The basic assumption of Kaluza and Klein was that the correct vacuum is the space $M^4\times S^1$, the product of four-dimensional Minkowski space with a circle of radius R. One might question whether this is a reasonable assumption. In fact were it not for every day experience one might expect the vacuum to be the more symmetric five-dimensional Minkowski space M^5. Since in general relativity one cannot compare the 'energy' of these two solutions the only criterion is stability. Classically both vacua are stable, however Witten has argued [8] that $M^4\times S^1$ is a "false vacuum" that undergoes quantum mechanical decay. (One could however stabilize the Kaluza–Klein vacuum by introducing fermion fields [8].) We shall ignore this defect of the five-dimensional theory, along with the other defects of this unrealistic theory, but one should be aware that the issue of stability is one that must be addressed in any version of a realistic Kaluza–Klein theory.

The radius of the circle in the fifth dimension, R, is undetermined by the classical equations of motion, since any circle is flat. If R is sufficiently small then all low-energy experiments will simply average over the fifth dimension. In fact the components of the metric, $g_{AB}(x^\mu, x^5)$, can be expanded in a Fourier series

$$g_{AB}(x^\mu, x^5)=\sum_{n=-\infty}^{+\infty} g_{AB}(x^\mu)\,\mathrm{e}^{inx^5/R}\,, \tag{2}$$

and all modes with $n\neq 0$ will have energies greater than hc/R. Thus the effective low-energy theory can be deduced by considering the metric g_{AB} to be independent of x^5. Under these conditions the theory is invariant under general coordinate transformations that are independent of x^5. In addition to ordinary four-dimensional coordinate transformations $x^\mu\to x^\mu(x^\nu)$, we have a U(1) local gauge transformation $x^5\to x^5+\Lambda(x^\mu)$, under which $g_{\mu 5}(x)$ transforms as a vector gauge field

$$g_{\mu 5}(x)\to g_{\mu 5}(x)+g_{55}(x)\partial_\mu\Lambda\,. \tag{3}$$

Therefore the low-energy theory should be a theory of four-dimensional gravity plus a U(1) gauge theory, i.e. electromagnetism, with the massless modes of $g_{\mu\nu}(g_{\mu 5})$ corresponding to the graviton (photon). The low-energy theory is also invariant under scale transformations of the fifth dimension

$$x^5+\lambda x^5\,, \qquad g_{55}\to\frac{1}{\lambda^2}g_{55}\,, \qquad g_{5\mu}\to\frac{1}{\lambda}g_{5\mu}\,. \tag{4}$$

This global scale invariance is spontaneously broken by the Kaluza–Klein vacuum (since R is fixed) thus giving rise to a Goldstone boson: the dilaton. This particle is actually a 'pseudo Goldstone boson', since dilatation invariance is broken once x^5 dependent quantum corrections are taken into account. It is conceivable, as argued recently by Appelquist and Chodos [9] that quantum corrections can give the dilaton a mass and fix the value of R. Ignoring this possibility R is determined by hand so as to yield the observed values of the electric charge and Newton's constant. In a more realistic theory, in which the internal space has non-vanishing curvature, the internal dimensions are dynamically determined, pseudo Goldstone bosons will not necessarily exist, and the gauge couplings will be calculable.

To exhibit the low-energy theory we write the metric as follows

$$g_{AB}=\begin{pmatrix} g_{\mu\nu}+A_\mu A_\nu V & A_\mu V \\ A_\nu V & V \end{pmatrix}, \qquad g^{AB}=\begin{pmatrix} g^{\mu\nu} & -A^\mu \\ -A^\nu & V^{-1}+A_\lambda A^\lambda \end{pmatrix},$$

$$g_5 \equiv \det(g_{AB})=\det(g_{\mu\nu})\cdot V \equiv g_4 V,$$

$$ds^2=V(dx^5+A_\mu\, dx^\mu)^2+g_{\mu\nu}\, dx^\mu\, dx^\nu . \tag{5}$$

The five-dimensional curvature scalar can be expressed in terms of the four dimension curvature R_4, the field strength $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\mu A_\mu$, and the scalar field V,

$$R_5=R_4-\frac{2}{V^{1/2}}\Box V^{1/2} . \tag{6}$$

Thus the effective low-energy theory is described, by the four-dimensional action

$$S=\frac{-1}{16\pi G}\int d^4x\,\sqrt{-g^4}\,V^{1/2}\{R_4+\tfrac{1}{4}VF_{\mu\nu}F^{\mu\nu}\}, \tag{7}$$

where we have dropped the terms in R_5 involving V, since these yield (when multiplied by $\sqrt{-g_5}$) a total derivative, and

$$G=\frac{G_{\mathrm{K}}}{2\pi R}, \tag{8}$$

is Newton's constant, which is determined by $\sqrt{\hbar G/c^3}=1.6\times 10^{-33}$ cm.

This theory is recognizable as a variant of the Brans–Dicke theory [10] of gravity, with $V^{1/2}$ identified as the Brans–Dicke massless scalar field, coupled to electromagnetism. V indeed sets the local scale of the gravitational coupling. In the vacuum $V=1$. Unlike Brans–Dicke theory, which contains a coupling $\omega\partial_\mu V\,\partial^\mu V/V^2$ with an arbitrary coupling constant ω, here $\omega=0$. (In the real world ω must be $\gtrsim 500$). Also in the Brans–Dicke theory the coupling of V to matter is somewhat arbitrary; here it is totally fixed by five-dimensional general covariance.

The radius R is determined by the electric charge. To see this consider a complex field ϕ with action: $S_\phi=\int d^5x\,\sqrt{-g^5}\,(\partial_A\phi)(\partial_B\phi^\dagger)g^{AB}$. The Fourier component of ϕ with non-trivial x^5 dependence, $\phi^{(n)}(x)\,e^{inx^5/R}$, will behave as a particle of charge $e=\sqrt{16\pi G}\,n/R$ and mass $=n/R$, since

$$(\partial_A\phi)(\partial_B\phi^\dagger)g^{AB}=\left|\left(\partial_\mu+i\frac{n}{R}A_\mu\right)\phi\right|^2+V^{-1}\frac{n^2}{R^2}|\phi|^2.$$

(Note that the properly normalized gauge field is $(16\pi G)^{-1/2}A_\mu$). Thus

$$\alpha=\frac{e^2}{4\pi\hbar c}=\frac{4\hbar G}{c^3R^2}, \qquad R=\frac{2}{\sqrt{\alpha}}\sqrt{\frac{\hbar G}{c^3}}\simeq 3.7\times 10^{-32}\text{ cm}. \tag{9}$$

As stated above, the coupling of matter to the four-dimensional Brans–Dicke like theory is fixed by five-dimensional general covariance. Consider a classical point test particle of mass m. It has action

$$S_m = m \int \mathrm{d}\tau \left(g_{AB} \frac{\mathrm{d}x^A}{\mathrm{d}\tau} \frac{\mathrm{d}x^B}{\mathrm{d}\tau} \right)^{1/2} , \tag{10}$$

and the motion is given by a five-dimensional geodesic

$$\frac{\mathrm{d}^2 x^A}{\mathrm{d}\tau^2} + \Gamma^A_{BC} \frac{\mathrm{d}x^B}{\mathrm{d}\tau} \frac{dx^C}{\mathrm{d}\tau} = 0 . \tag{11}$$

Now since the space–time possesses a Killing vector,

$$K = K^A \frac{\partial}{\partial x^A} = \frac{\partial}{\partial x^5} ,$$

it is guaranteed that $K_A\, \mathrm{d}x^A/\mathrm{d}\tau$ is a constant of the motion. Indeed a first integral of eq. (11) is

$$K_A \frac{\mathrm{d}x^A}{\mathrm{d}\tau} = V \frac{\mathrm{d}x^5}{\mathrm{d}\tau} + A_\mu \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \equiv \frac{q}{m} . \tag{12}$$

The remaining equations of motion then take the form

$$m \left(\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau} \frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} \right) = q F^\mu{}_\nu \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} + \frac{q^2}{2m} \frac{\partial^\mu V}{V^2} , \tag{13}$$

($\Gamma^\mu_{\alpha\beta}$ is the four-dimensional connection, constructed from $g_{\mu\nu}$.) We recognize on the right the Lorentz force if $q\sqrt{16\pi G}$ is identified with the charge of the particle, plus an interaction with the scalar field V.

As is customary in Brans–Dicke theories [11], it is useful to recast the effective low-energy theory in a more familiar form, in which the coefficient of $\sqrt{-g_4}\, R$ is constant. This can be achieved by a conformal transformation under which $g_{\mu\nu}$ and V are both scaled by a factor equal to $V^{-1/3}$. ((Recall that when

$$g_{\mu\nu} \to \lambda^{-1} g_{\mu\nu} , \qquad R_4 \to \lambda \left(R_4 + 3 \Box \ln \lambda - \frac{3}{2} \left(\frac{\partial_\mu \lambda\, \partial^\mu \lambda}{\lambda^2} \right) \right) .$$

Thus if instead of eq. (5) we take (to avoid confusion we replace V by σ when using this form of g_{AB})

$$g_{AB} = \sigma^{-1/3} \begin{pmatrix} g_{\mu\nu} + A_\mu A_\nu \sigma & A_\mu \sigma \\ A_\nu \sigma & \sigma \end{pmatrix} , \tag{14}$$

the low-energy action assumes the form:

$$S = -\frac{1}{16\pi G} \int \mathrm{d}^4 x \sqrt{-g^4} \left[R_4 + \tfrac{1}{6} g_{\mu\nu} \frac{\partial_\mu \sigma\, \partial_\nu \sigma}{\sigma^2} + \tfrac{1}{4} \sigma F_{\mu\nu} F^{\mu\nu} \right] . \tag{15}$$

We can therefore consider the Kaluza–Klein theory at low energies to describe a massless scalar, $\psi = \frac{1}{3}\ln\sigma$, coupled to gravity and to electromagnetism. The gravitational coupling is σ independent, however the local strength of the electromagnetic coupling is proportional to $\sqrt{\sigma}$. Also the mass of the other matter fields will depend on the local value of σ, since under this conformal transformation the four-dimensional mass will transform as

$$m \to \bar{m} = m\sigma^{-1/6} . \tag{16}$$

We shall see below that the unusual gravitational properties of our solitons are due to this coupling of the dilaton to matter.

3. The Kaluza–Klein monopole

We now turn to the construction of solitons in Kaluza–Klein theories. By solitons we mean non-singular solutions of the classical field equations which represent spatially localized lumps that are topologically stable. The paradigm of such a soliton is the magnetic monopole of 't Hooft and Polyakov [12] that occurs in non-abelian gauge theories. Such solutions are expected to have a large mass (of order M/g, where M is the mass scale of the theory and g a dimensionless coupling), and are the starting point for the semiclassical construction of quantum mechanical particle states.

Our goal is to construct solutions of the five-dimensional field equations that approach the vacuum solution: $g_{AB} = \eta_{AB}$ (or $V = 1$, $A_\mu = 0$, $g_{\mu\nu} = \eta_{\mu\nu}$) at spatial infinity. For a variety of topological reasons one has an *a priori* reason to expect these to occur. Abelian gauge theories allow for the existence of Dirac monopoles. Kaluza–Klein theory contains electromagnetism; thus one expects that it allows for solutions with magnetic change. Our theory can be viewed as a principal fiber bundle with M^4 as the base manifold and U(1) as the structure group. The vacuum is the trivial bundle $M^4 \times S^1$; however there exist topologically inequivalent bundles. At spatial infinity a solution will describe an S^1 bundle over S^2 (the boundary of 3-space), and there exist an infinite collection of such bundles, each characterized by an integer, which will be identical to the magnetic charge of the solution.

Of course it is not *a priori* obvious that there exist regular magnetic monopole solutions. To search for such solutions it is natural to consider *static*, metrics with $\partial/\partial t$ as a Killing vector. We therefore assume $(\partial/\partial t)g_{AB} = 0$. It is also natural to look for solutions with $g_{0A} = \delta_{0A}$. This is analogous to constructing the Yang–Mills monopole as a static solution in $A_0 = 0$ gauge, which does yield the minimum mass monopole. In this case the space-time is totally flat in the "time" direction and the field equations are simply

$$R_{ij} = R_{5i} = R_{55} = 0 , \tag{17}$$

namely the four-dimensional, wholly spacelike, manifold at each fixed t has vanishing Ricci tensor. These are simply the equations of four-dimensional euclidean gravity, where we can think of x^5 are representing euclidean, periodic, time.

In conventional, four-dimensional, relativity theory our search would end here. If $R_{ij}=0$ then $R^i{}_{jnl}=0$ and the three manifold is flat. Thus there exist no static solitons in ordinary gravity. This is no longer true once the number of spatial dimensions is greater than three. We therefore expect that in all Kaluza–Klein theories, static soliton solutions will exist as dictated by topology.

Our task of finding solutions to eq. (17) is greatly simplified by the fact that these equations are those of euclidean gravity in four dimensions and have been extensively studied. Many solutions are known, including many with the correct global structure. We will, in fact, utilize many well known solutions (including the Taub–NUT, Kerr and Schwarzchild metrics) to construct our solitons.

The simplest and basic soliton is the magnetic monopole. It is a generalization of the self-dual euclidean Taub–NUT solution [13], and is described by the following metric

$$(\mathrm{d}s)^2=-\mathrm{d}t^2+V(\mathrm{d}x^5+4m(1-\cos\theta)\,\mathrm{d}\phi)^2+\frac{1}{V}(\mathrm{d}r^2+r^2\,\mathrm{d}\theta^2+r^2\sin^2\theta\,\mathrm{d}\phi^2)\,,$$

$$\frac{1}{V}=1+\frac{4m}{r}\,, \tag{18}$$

where (r,θ,ϕ) are polar coordinates. Setting $\mathrm{d}t=0$, the resulting subspace is the Taub–NUT instanton. We are therefore assured that this is a solution of the five-dimensional equations. There would appear to be a coordinate singularity when $r=0$, where the Killing vector field has a fixed point. This is a so called NUT singularity and is absent if x^5 is periodic with period $16\pi m$ [14]. Since we want our solution to approach the vacuum for large r we must identify $16\pi m$ with $2\pi R$. Thus

$$m=\tfrac{1}{8}R=\frac{\sqrt{\pi G}}{2e}\,. \tag{19}$$

The gauge field, A_μ, is clearly that of a monopole

$$A_\phi=4m(1-\cos\theta)\,,$$

$$\boldsymbol{B}=\boldsymbol{\nabla}\times\boldsymbol{A}=\frac{4m\boldsymbol{r}}{r^3}\,. \tag{20}$$

and has a Dirac string singularity running from $r=0$ to ∞. As usual this singularity is an artifact if and only if the period of x^5 is equal to that of $16\pi m$. This is simply the geometrical analogue of Dirac's argument. The metric is regular on the half axis $\theta=0$, but has a singularity at $\theta=\pi$ since the $(1-\cos\theta)$ term in the metric means that a small loop about this axis does not shrink to zero length at $\theta=\pi$. By a change of coordinates $x^{5\prime}=x^5+8m\phi$ the metric becomes regular at $\theta=\pi$ but not at $\theta = 0$. One can then use $(x^5, t, r, \theta, \phi)$ to cover the northern hemisphere, and $(x^{5\prime}, t, r, \theta, \phi)$ to cover the southern hemisphere. Since x^5 and ϕ are periodic with periods $2\pi R$ and 2π respectively, one must have that $2\pi R = 16\pi m$.

The magnetic charge of our monopole is thus fixed by the radius of the Kaluza–Klein circle. If we scale the magnetic field so as to have the proper normalization, $B \to (16\pi G)^{-1/2} B$, we find that the magnetic charge g is (using eq. (9))

$$g = \frac{4m}{\sqrt{16\pi G}} = \frac{R}{2\sqrt{16\pi G}} = \frac{1}{2e}\,. \tag{21}$$

Thus, as expected, our monopole has one unit of Dirac charge.

What is the mass of this soliton? A gravitational field possesses a conserved energy-momentum if it is asymptotically flat. The five-dimensional energy momentum pseudo-tensor is given by

$$(-g_5)t^{AB} = \partial_C h^{ABC}\,, \qquad h^{ABC} = \frac{1}{16\pi G_{\mathrm{K}}}\,\partial_D[-g_5(g^{AB}g^{CD} - g^{AC}g^{BD})]\,. \tag{22}$$

The conserved five-momentum $P^A = \int \mathrm{d}^3x\, \mathrm{d}x^5\, (-g_5)\, t^{A0}$ is generally covariant if $g_{AB} \to \eta_{AB} + \mathrm{O}(1/r)$ as $r \to \infty$, and can be written in terms of the components of g_{AB} at spatial infinity

$$P^A = \frac{1}{16\pi G_5}\int \mathrm{d}^3x\, \mathrm{d}x^5\, \partial_i\, \partial_B[-g_5(g^{A0}g^{iB} - g^{Ai}g^{0B})]\,. \tag{23}$$

The fifth component of P^A is simply the conserved electric charge of the field as can be seen by setting $A = 5$. Since the integrand is a total spatial divergence, it can be evaluated as $r \to \infty$ where it is given by (the index $b \neq 5$ since g^{AB} is periodic in x^5)

$$\partial_i\partial_j(g^{50}g^{ij} - g^{5i}g^{0j}) \underset{r\to\infty}{\simeq} \partial_i[\partial_0 A_i - \partial_i A_0] = \boldsymbol{\nabla}\cdot\boldsymbol{E}\,.$$

For our soliton, $P^A = M\delta^{A0}$, where M is the mass of the soliton and

$$M = -\frac{2\pi R}{16\pi G_{\mathrm{K}}}\int \mathrm{d}^3x\, \nabla^2 \frac{1}{V} = \frac{m}{G}\,. \tag{24}$$

Since m is fixed by the radius of the vacuum circle, which in turn is fixed by e and G, the soliton mass is determined to be

$$M^2 = m_{\mathrm{P}}^2/16\alpha\,, \tag{25}$$

where $m_{\mathrm{P}} = \sqrt{\hbar c/G} = 2.17\cdot 10^{-5}$ g is the Planck mass. As is customary for solitons, the monopole mass is $1/e$ times heavier than the mass scale of the theory. Since the soliton solution is asymptotically flat, we can easily construct moving soliton solutions, by applying Lorentz boosts to the soliton at rest. Since these preserve the boundary conditions

$$g'_{AB}(x) = \Lambda_A{}^{B'} \Lambda_A{}^{B'} g_{A'B'}(\Lambda x) \,,$$

where Λ_{AB} is a four-dimensional Lorentz transformation ($\Lambda_{A5} = \delta_{A5}$), will yield a new solution, corresponding to a monopole moving with constant velocity. It will have an energy-momentum, P_μ, with $P_\mu P^\mu = M^2$.

These solitons can be quantized by using methods that are familiar in ordinary gauge theories [15] and will yield (ignoring the non-renormalizability of the radiative corrections) particles of mass $M(1+\mathrm{O}(\alpha))$.

In Yang–Mills theories one can sometimes find multi-monopole static solutions (the Prasad–Sommerfeld limit), by arranging for the electromagnetic repulsion of the monopoles to cancel with an attractive force due to scalar Higgs exchange. It is a remarkable property of Kaluza–Klein monopoles that they do not interact. In fact one can easily construct multi-monopole solutions [16], with a similar line element to eq. (18)

$$(\mathrm{d}s)^2 = -\mathrm{d}t^2 + V[\mathrm{d}x^5 + A_i\, \mathrm{d}x^i]^2 + \frac{1}{V}(\mathrm{d}r^2 + r^2\, \mathrm{d}\Omega^2) \,,$$

$$\frac{1}{V} = 1 + \sum_{i=1}^{N} \frac{4m}{|\boldsymbol{x} - \boldsymbol{x}^i|} \,, \tag{26}$$

$$\partial_i \left(\frac{1}{V} \right) \equiv -\varepsilon_{ijk}\, \partial_j A_k = B_i \,. \tag{27}$$

This defines a regular manifold, with N solitons sitting at rest at $\boldsymbol{x} = \boldsymbol{x}_i$. As long as $m = \frac{1}{8}R$, there is no singularity at $\boldsymbol{x} = 0$ or $\boldsymbol{x} = \boldsymbol{x}_i$. This clearly describes a static configuration of N monopoles. The inertial mass of this configuration is clearly equal to NM. Thus the monopoles do not interact. In sect. 5 we shall explain why this occurs. We can also construct solitons with negative magnetic charge anti-monopoles, simply by taking $\partial_i(1/V) = +\varepsilon_{ijk}\partial_j A_k$, in eq. (27).

4. Dipole solitons

In ordinary Yang–Mills theories one does not find solitons which are magnetic dipoles. The reason is that a superposition of a monopole and an anti-monopole has the topology of the vacuum. Consequently, there is nothing to prevent these from annihilating each other. The situation is quite different in Kaluza–Klein theories. Static solitons can be classified according to the topology of their spatial subspaces. Thus the solitons of the Kaluza–Klein theory can be labelled by the Euler character χ and the Hirzebruch signature τ of the $t = \text{const}$ surfaces. It is easy to calculate χ and τ from a knowledge of the sources of the electromagnetic field. The only sources of the electromagnetic field are the fixed points of the Killing vector $\partial/\partial x^5$ where $V = 0$. As remarked earlier, $\partial/\partial x^5$ is a Killing vector which generates the U(1) symmetry of the theory. V is the norm of $\partial/\partial x^5$ which vanishes only at these fixed points. These fixed points may be either two-dimensional, "bolts", or zero-dimensional, "nuts". If the fixed point is a bolt it can be regarded as a two-dimensional space. There are two quantities which describe this situation. The first intrinsic, the Euler character, χ, of the two-dimensional surface. This takes the value two for a sphere, zero for torus and negative for surfaces of higher genus. The second is extrinsic, describing the embedding in the four-dimensional space: the self-intersection number, Y. The self-intersection number describes how the bolt behaves under infinitesimal deformations.

Nuts are generally described by a pair of co-prime integers p, q. In our cases, it turns out that at the nuts the curvature is either self-dual so that $p=q=1$ or anti self-dual when $p=-q=1$. Finally, the boundary of the space-time at spatial infinity will be an S^1 bundle over S^2 with k twists in it, where k is the monopole charge in Dirac units. One can now use the G-index theorems [17] to calculate χ and τ given this information.

$$\chi = \sum_{\substack{\text{nuts}\\ \text{antinuts}}} 1 + \sum_{\text{bolts}} \chi \,, \tag{28}$$

$$\tau = \sum_{\text{bolts}} Y + k - \varepsilon \,, \tag{29}$$

where $\varepsilon=0$ if $k=0$ and $\varepsilon=1$ if $k\neq 0$.

For the case of the vacuum, $\partial/\partial x^5$ has no fixed points and $\varepsilon=0$. Thus $\chi=\tau=0$. For the Kerr–Taub–NUT solutions, each nut is either self-dual (for monopoles) so that $\varepsilon=1$, or antiself-dual (for antipoles), with $\varepsilon=-1$, thus $\chi=N$ and $\tau=\pm(N-1)$. Thus, one expects that a monopole plus anti-monopole will be in a topological sector with $\chi=2$, and $\tau=0$, which is, of course, distinct from the vacuum. Since one cannot, at least by smooth classical evolution, deform this space to one with $\chi=0$ and $\tau=0$, the monopole anti-monopole pair will not classically annihilate. Thus, we expect to be able to construct soliton solutions with zero overall magnetic charge, but non-vanishing dipole moment.

Once again it is easy to find these solitons, we can simply reinterpret some well known solutions of euclidean gravity in four dimensions. In fact the Kerr–Schwarzschild metrics, which are regular solutions for $3+1$ spatial dimension in which the fourth dimension is periodic, will describe Kaluza–Klein dipoles. The line element is [18]

$$(\mathrm{d}s)^2 = -\mathrm{d}t^2 + \frac{1}{r^2 - a^2\cos^2\theta}[\Delta(\mathrm{d}x^5 + a\sin^2\theta\,\mathrm{d}\psi)^2 + \sin^2\theta((r^2-a^2)\,\mathrm{d}\psi - a\,\mathrm{d}x^5)^2]$$
$$+ (r^2 - a^2\cos^2\theta)\left[\frac{\mathrm{d}r^2}{\Delta} + \mathrm{d}\theta^2\right], \tag{30}$$

where $\Delta = r^2 - 2mr - a^2$. This manifold is regular for $r \geqslant m + \sqrt{m^2+a^2}$ (so that

$\Delta \geq 0$), $0 \leq \theta \leq \pi$, $0 \leq \psi \leq 2\pi$, $0 \leq \chi^5 \leq 2\pi/k$, if x^5 is identified with period $2\pi/\kappa$ and $\phi = \psi - \Omega x^5$ is identified with period 2π. Here

$$\kappa = \frac{\sqrt{m^2+a^2}}{2m[m+\sqrt{m^2+a^2}]}, \qquad \Omega = \kappa \frac{a}{\sqrt{m^2+a^2}}.$$

As $r \to \infty$ the metric becomes flat, and the global structure is identical to that of the Kaluza–Klein vacuum if we take (r, θ, ϕ) to be polar coordinates of ordinary 3-space and identify $1/\kappa$ with R.

This solution does describe a magnetic dipole. The vector potential is given, for large r, by

$$A_\phi \underset{r\to\infty}{\sim} \frac{-2ma \sin^2\theta}{r}. \tag{31}$$

This is precisely the field of a magnetic dipole, located at the origin, of strength $2ma$, pointing along the z axis. It corresponds, we argue, to the stable configuration one obtains by combining monopole and anti-monopole solitons [19]. The Euler character of this solution is 2 and the signature is 0, consistent with this interpretation. This dipole is not produced by rotating currents and this soliton has zero angular momentum. It is best pictured as a monopole anti-monopole pair separated by a distance proportional to a. Therefore we would expect that the mass of the dipole should approach twice the monopole mass as a becomes large.

To evaluate the mass, M, of the soliton we examine the asymptotic form of the metric (as $r \to \infty$), which is independent of a, and the same as the Schwarzschild metric

$$(\mathrm{d}s)^2 \underset{r\to\infty}{\approx} (\mathrm{d}s)^2_{\text{Schw.}} = -(\mathrm{d}t)^2 + \frac{\mathrm{d}r^2}{1-2m/r} + r^2\,\mathrm{d}\Omega^2 + \left(1-\frac{2m}{r}\right)\mathrm{d}x_5{}^2. \tag{32}$$

Using eq. (23) we find that

$$M = \tfrac{1}{2}m. \tag{33}$$

But the period of x^5, $2\pi/\kappa$, must be identified with $2\pi R$, so

$$R = 2m\frac{m+\sqrt{m^2+a^2}}{\sqrt{m^2+a^2}}. \tag{34}$$

Therefore for a given value of a the dipole mass is determined. When $a = \infty$, corresponding to infinite dipole moment, $R(a=\infty) = 2m$ and $M(a=\infty) = \frac{1}{4}R =$ 2 × monopole mass. As we let the monopole anti-monopole approach each other, a will decrease and the dipole will become lighter. The minimum dipole mass is obtained when $a = 0$, where $R(a=0) = 4m$ and $M(a=0) = \frac{1}{8}R =$ monopole mass.

This solution has only one fixed point of the Killing vector $\partial/\partial x^5$, which is at the source of the magnetic dipole, $r = m + (m^2+a^2)^{1/2}$. This surface is a spherical bolt with $Y = 0$. The boundary at infinity is a direct product of $\mathrm{S}^1 \times \mathrm{S}^2$, thus $q = 0$. Hence as expected the G-index theorems show that $\chi = 2$ and $\tau = 0$.

We expect to find representatives of most topological classes amongst the solutions to the five-dimensional field equations. Thus, for example we might expect to find solutions which contain both elementary monopoles and elementary dipoles. We can explicitly exhibit such a solution which contains a single elementary monopole and a single dipole. It too is related to a gravitational instanton, namely the Kerr–Taub–Bolt solutions [20]. Like the previous solutions, it is not associated with any space-time singularity. The line element is given by

$$ds^2 = -dt^2 + \frac{1}{\Xi}[\Delta(dx^5 + P\,d\psi)^2 + \sin^2\theta(a\,dx^5 + Q\,d\psi^2)] + \Xi\left(\frac{dr^2}{\Delta} + d\theta^2\right), \tag{35}$$

where

$$\Delta = r^2 - 2mr + N^2 - a^2,$$

$$\Xi = r^2 - (N + a\cos\theta)^2,$$

$$P = a\sin^2\theta + 2N\cos\theta - \frac{aN^2}{N^2 - a^2},$$

$$Q = r^2 - a^2 - \frac{N^2}{N^2 - a^2}.$$

In the limit $N = 0$, we obtain the Kerr–Schwarzschild metrics described earlier. In the limit $N = \pm m$, $m > 0$, $a = 0$, we obtain the Taub–NUT metrics described in sect. 3. However, we can consider the general case with fixed non-zero $N \neq \pm m$, in which case we find a coordinate singularity at $r = r_+ = m + (m^2 - N^2 - a^2)^{1/2}$ where Δ vanishes. This singularity is a bolt, and there is no spacetime singularity if x^5 is identified with period $8\pi|N|$, and

$$4|N| = \frac{r_+^2 - a^2 - N^2/(N^2 - a^2)}{(m^2 - N^2 - a^2)^{1/2}}. \tag{36}$$

Thus the radius of the Kaluza–Klein circle R is given by

$$R = 4|N| = \frac{4\sqrt{\pi G}}{e}. \tag{37}$$

Provided that $\Xi > 0$ for $r > r_+$, this space will contain no curvature singularities. One can solve these above conditions which then show that for fixed N, one can find m as a function of a. For details of this procedure, see ref. [20]. This results in two distinct families, one with

$$0 \leqslant a \leqslant 0.693|N|,$$

and the other with

$$\tfrac{1}{4}(1 + \sqrt{17})|N| \leqslant a < \infty.$$

One can easily calculate the vector potential for large r, which gives

$$A_\phi \xrightarrow{r\to\infty} \left[2N\cos\theta - \frac{2aN^2}{N^2-a^2} + \frac{2m}{r}\left(a\sin^2\theta - 2N\cos\theta + \frac{aN^2}{N^2-a^2}\right)\right] + \mathrm{O}(r^{-2})\,. \quad (38)$$

This is the vector potential of a magnetic monopole of moment $2N$ superimposed of a magnetic dipole of moment $2ma$. The fixed point of $r = r_+$ is a spherical bolt with self-intersection number $Y = -1$. The boundary at infinity is not a direct product, thus $\epsilon = 1$, and so $\chi = 2$ and $\tau = 1$. We can evaluate the inertial mass of this object which turns out to be $M = \frac{1}{2}m$.

Two limits are easy to understand. Firstly if $a = 0$, then $m = \frac{5}{4}|N|$ so that the mass of the soliton becomes $M = \frac{5}{8}|N| = \frac{5}{32}R$. Secondly, if $a \to \infty$ $m \to 2|N|$, thus the mass of the soliton becomes $M = |N| = \frac{1}{4}R$ which is exactly what happens in the Kerr case. The energetics of this are thus very similar to the example of Kerr discussed earlier in this section.

5. Gravitational properties of the solitons

In this section we explore the gravitational properties of our solitons and their interactions. These turn out to be rather strange, contradicting naive interpretations of the principle of equivalence. Thus we shall show that although the solitons possess inertial mass their "gravitational mass" is zero. We shall explain the reasons for this unusual phenomenon in both the five-dimensional and the effective four-dimensional formulations of Kaluza–Klein theory.

The strange gravitational properties of the solitons is clearly visible in the form of the five-dimensional metric, g_{AB}. All of the soliton solutions have been constructed as four-dimensional, euclidean, solutions, with the time coordinate grafted on. They are all completely "flat" in the time direction, they all have $g_{A0} = \delta_{A0}$, $(\partial/\partial t)g_{AB} = 0$. Therefore they all possess time like geodesics corresponding to test particles sitting at rest relative to the soliton. For each of the solutions $x^0 = \tau$, $\boldsymbol{x} = \boldsymbol{x}_0$, $x^5 = {x_0}^5$ is a geodesic. This describes the world line of a neutral $(\mathrm{d}x^5/\mathrm{d}\tau = 0)$ test particle, which remains at rest with respect to the soliton, experiencing no force. The newtonian force, that a slowly moving test particle experiences, which is proportional to $\frac{1}{2}\boldsymbol{\nabla} g_{00}$, vanishes.

This turn of events violates our usual, four-dimensional, understanding of the principle of equivalence. Our solitons certainly have inertial mass, however they do not exert a newtonian force on a test particle at rest. According to the usual definition their gravitational mass is zero. They will, of course, interact with some test particles. If the particles are charged $(\mathrm{d}x_5/\mathrm{d}\tau \neq 0)$ they will experience velocity dependent (electromagnetic) forces, and even the totally neutral soliton will exert velocity dependent forces on moving test particles. For example one can easily verify that massless particles are deflected, at large distances from the soliton, as if they were ordinary objects whose gravitational mass was equal to their inertial mass. What is strange however that a neutral observer at rest cannot tell whether a soliton is in his vicinity, since he experiences no force. However this is not a violation of the principle of equivalence, since we have not violated general covariance in five dimensions. Test particles do move along geodesics, it is just that in the vicinity of a massive soliton there exist static geodesics.*

What is new in the five-dimensional theory is the fact that static geodesics ($x^A = x_0^A + \delta_0^A \tau$) exist for both the vacuum and the soliton background metrics. *This is a violation, not of the principle of equivalence, but of Birkhoff's theorem.* In ordinary four-dimensional relativity, Birkhoff's theorem states that the Schwarzschild metric is the unique spherically symmetric solution of the field equations. Thus far away from any local lump the metric is

$$(\mathrm{d}s)^2 \underset{r\to\infty}{\approx} -\left(1-\frac{2M}{r}\right)(\mathrm{d}t)^2 + \frac{\mathrm{d}r^2}{1-2M/r} + r^2\,\mathrm{d}\Omega^2\,,$$

which means the inertial mass in M (according to eq. (21)) and the gravitational potential is $\frac{1}{2}(g_{00}+1) = M/R$. In the Kaluza–Klein theory Birkhoff's theorem is no longer true. In fact we already know of two different spherically symmetric solutions: the four-dimensional Schwarzschild metric with $g_{5A} = \delta_{5A}$, and the zero dipole moment soliton (eq. (32)). In fact one can construct a 2-parameter family of spherically symmetric solutions of five-dimensional relativity. They are described by the line element

$$(\mathrm{d}s)^2 = -\left(\frac{1-m/r}{1+m/r}\right)^{2/\alpha}(\mathrm{d}t)^2 + \left(1+\frac{m}{r}\right)^4\left(\frac{1-m/r}{1+m/r}\right)^{2(\alpha-\beta-1)/\alpha}(\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2)$$
$$+\left(\frac{1-m/r}{1+m/r}\right)^{2\beta/\alpha}(\mathrm{d}x^5)^2\,, \tag{39}$$

where $\alpha = \sqrt{\beta^2+\beta+1}$. This solution interpolates between the Schwarzschild metric, $\beta = 0$, $\alpha = 1$, and the soliton solution $\beta = \alpha = \infty$. It is only non-singular for the soliton case (when x^5 is periodic), the Schwarzschild solution is singular at $r = 0$, the others are singular at $r = m$.

In Kaluza–Klein theory the above line element would describe the exterior geometry of a star. The inertial mass of the star can be determined by the asymptotic behavior of eq. (39), assuming no interior singularities, and is equal to $M_{\mathrm{IN}} = (1+\beta)m/\alpha$. The ratio of β to α must be determined by the form of the five-dimensional energy momentum tensor T_{AB} which is non-vanishing in the stars' exterior. In particular if $T_{AB} = 0$ then $\beta = \alpha = \infty$. In general $\beta \neq 0$. The "gravitational mass" of the stars can be determined by the asymptotic form of g_{00}, and is given by $M_{\mathrm{grav}} = m/\alpha$. Thus only for $\beta = 0$ are these two masses equal.

The non-uniqueness of the exterior geometry of a massive object in Kaluza–Klein theory raises an interesting question as to the final fate of collapsing matter. There

* Note that the soliton itself cannot be treated as a test particle since it cannot be deformed smoothly into an infinitesimal perturbation of flat space, as is clear from the fact that $GM/R \sim 1$ for the soliton.

appear to be two possibilities: a collapsing star could end up described by the Schwarzschild metric, i.e. a black hole, or it might end up as a soliton, a totally non-singular metric. Can one thereby avoid, in Kaluza–Klein theories, gravitational collapse to a black hole? This possibility deserves to be explored, however the strange gravitational properties of Kaluza–Klein theories are intimately connected with one of its least attractive features – the existence of the pseudo-Goldstone massless scalar field, and presumably would be absent in a realistic theory which had no such particles.

It is very instructive to analyze the interaction of the solitons from the point of view of the effective four-dimensional theory. We saw, in sect. 2, that the Kaluza–Klein theory could be viewed at low energies as the theory of a massless scalar field coupled to gravity and to electromagnetism. The soliton solutions we have constructed are soliton solutions of this four-dimensional theory as well. Consider for example the zero dipole moment soliton. If we use the form of eq. (14) for the metric this describes a four-metric $g_{\mu\nu}$ and a massless scalar field ψ, with

$$(\mathrm{d}s)^2 = -\sqrt{1-\frac{2m}{r}}\,(\mathrm{d}t)^2 + \frac{\mathrm{d}r^2}{\sqrt{1-2m/r}} + \sqrt{1-\frac{2m}{r}}\,r^2\,\mathrm{d}\Omega^2\,,$$

$$\psi = \tfrac{1}{2}\ln\left(1-\frac{2m}{r}\right) = \tfrac{1}{3}\ln\sigma\,, \qquad \infty \geqslant r \geqslant 2m\,. \tag{40}$$

Although this is a singular solution of the effective four-dimensional theory it is a perfectly sensible soliton. The singularity arises because the conformal factor, $\sigma^{-1/3}$, is singular at $r = 2m$.

Now let us consider a massive test particle and its interaction with this soliton. The test particle will experience a gravitational force. In fact since $g_{00,r\to\infty} \approx 1 + m/r$, one might conclude that the "gravitational mass" of the soliton is $\frac{1}{2}m$ (equal to its inertial mass). However it also interacts with the scalar field, in a way dictated by five-dimensional general covariance. As noted in sect. 2, a test particle of mass μ in five dimensions, appears in four dimensions to have mass $\mu = \mu\sigma^{-1/6}$. Therefore the equation of motion is

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = U^\mu\,, \qquad \frac{\mathrm{d}}{\mathrm{d}\tau}(\bar{\mu} g_{\mu\nu} U^\nu) - \tfrac{1}{2}\bar{\mu}(\partial_\mu \gamma_{\alpha\beta}) U^\alpha U^\beta + \partial_\mu \bar{\mu} = 0\,. \tag{41}$$

It is easily verified that $U^\mu = \delta^\mu{}_0(-g_{00})^{-1/2} = \delta^\mu{}_0\sigma^{-1/6}$ is a solution of this equation. In other words the test particle can remain relatively at rest with respect to the soliton since the newtonian force, $\frac{1}{2}\bar{\mu}\,\partial_i(-g_{00})U^0U^0 = \frac{1}{2}\mu\sigma^{-1/2}\partial_i\sigma^{1/3}$, is exactly cancelled by the interaction with the scalar field, $\partial_i\bar{\mu} = \mu\,\partial_i\sigma^{-1/6}$. The "mass" of the test particle,

$$\bar{\mu} = \frac{\mu}{(1-2m/r)^{1/4}}\,,$$

increases in the vicinity of the soliton, giving rise to a repulsive force which precisely cancels the newtonian attraction.

This situation is familiar in Brans–Dicke theories [10], where particles do not move, in general, along geodesics. Here however, the deviation from geodesic motion is extreme, since the interaction with the Brans–Dicke field is totally fixed by five-dimensional covariance and must reproduce our five-dimensional analysis.

We should point out however that some objects will be attracted to the soliton. For example we have done a calculation that shows that an ordinary four-dimensional star (described by an exterior four-dimensional Schwarzschild solution with $g_{A5}=\delta_{A5}$) is attracted to the solitons with the usual newtonian attraction. The calculation consists of trying to construct a solution of the field equations which corresponds to a superposition of a soliton (in fact we considered the dipole moment zero soliton) and a Schwarzschild star separated by distance R. A lengthy calculation, which we will spare the reader, shows that such a solution can be constructed (for large R), however it is singular along a strut connecting the centers of the star and the soliton. This strut corresponds to a constraint that must be introduced, and is equivalent to the existence of an interaction energy between the star and the soliton which turns on to equal the newtonian potential $GM_{\text{star}}M_{\text{soliton}}/R$. This is not surprising. Such a star, which is we recall not the general form of a star in Kaluza–Klein theory, does not couple to the σ-field, and thus its newtonian attraction to the soliton is not cancelled. Thus the interaction of a given object with the soliton is not dictated only by its mass and its charge, but by what it is made of, which dictates its coupling to the scalar field.

In sect. 3 we noted that we could construct a multi-monopole soliton, eqs. (23), (24), which described a collection of non-interacting monopoles. One might have thought that the reason that they were non-interacting was their gravitational attraction was cancelled by their electromagnetic repulsion. However the situation is quite different, since the gravitational attraction is cancelled by the interaction with the scalar field. The electromagnetic interaction of the monopoles in fact vanishes. To see this consider the multi-monopole solution, given by eqs. (26), (27), but transformed to the form of eq. (14). Thus, the four-dimensional line element is

$$(\mathrm{d}s)^2=-\sqrt{V}(\mathrm{d}t)^2+\frac{1}{\sqrt{V}}(\mathrm{d}r^2+r^2\,\mathrm{d}\Omega^2)\,,$$

$$\frac{1}{V}=1+\sum_{i=1}^{N}\frac{4m}{|\boldsymbol{x}-\boldsymbol{x}_i|}\,, \tag{42}$$

and the scalar field and vector potential are given by

$$\sigma=V^{3/2}\,,\qquad \partial_i\left(\frac{1}{V}\right)=-\varepsilon_{ijk}\partial_jA_k\,. \tag{43}$$

The electromagnetic energy of this configuration is given by

$$\begin{aligned} \varepsilon &= \frac{1}{16\pi G}\int \mathrm{d}^3x\,\sqrt{-g_4}\,\sigma\tfrac{1}{4}F_{ij}F_{kl}g^{ik}g^{jl} \\ &= \frac{1}{16\pi G}\int \mathrm{d}^3x\,\tfrac{1}{2}V^2\partial_i\left(\frac{1}{V}\right)\partial_i\left(\frac{1}{V}\right). \end{aligned}$$

This expression is finite, since V vanishes at the poles of $1/V$. Now

$$V^2\partial_i\left(\frac{1}{V}\right)\partial_i\left(\frac{1}{V}\right) = -\partial_i V\partial_i\frac{1}{V} = -\partial_i\left(V\partial_i\frac{1}{V}\right) + V\nabla^2\frac{1}{V}, \qquad V\nabla^2\frac{1}{V} = 0.$$

Therefore,

$$\varepsilon = \frac{1}{32\pi G}\int \mathrm{d}s^i\cdot V\partial_i\frac{1}{V} = \frac{mN}{2G}. \tag{44}$$

Thus the electromagnetic energy of the collection of monopoles is independent of their separation (and equal to one half the inertial mass). This occurs because the effective charge is proportional to $1/\sqrt{\sigma}$ (in other words the action is $\sigma F_{\mu\nu}F^{\mu\nu}$) and therefore the self-energy of one monopole is smaller when it is close to another, exactly cancelling the increase in energy due to their repulsion.

These strange properties of Kaluza–Klein solitons are quite amusing, and they clarify the meaning of the principle of equivalence. However they are unlikely to persist in a realistic Kaluza–Klein theory. In the real world a Brans–Dicke scalar is almost excluded, certainly one that couples in the above fashion. In a realistic Kaluza–Klein theory such scalars will not be present, either because the internal dimensions are curved or because quantum corrections give the scalar a mass [9]. In that case there will be no long-range scalar interaction, and the solitons should exert ordinary gravitational forces. Thus if we observe a magnetic monopole with mass of order the Planck mass we expect it to fall.

6. Conclusions

The discovery of magnetic monopoles would provide strong evidence for unified gauge theories and allow one to probe the physics of unification. The monopole mass would, by itself, yield an estimate of the unification scale. A monopole whose mass is 10^{15} GeV would provide evidence for the SU(5) unification scheme. However, if the mass turns out to be of order the Planck mass, 10^{19} GeV, this would suggest a gravitational unification, perhaps along the lines of Kaluza–Klein theory.

The most important conclusion of this paper is that Kaluza–Klein theories can contain magnetic monopole solitons. The five-dimensional Kaluza–Klein theory that we have analyzed is highly unrealistic, however it does provide the simplest example of a magnetic monopole in a geometrically unified theory of gravity and electromagnetism. Their existence leads us to conjecture that Kaluza–Klein theories will, in general, possess regular classical soliton solutions whenever topological stability is indicated.

We can easily envisage solitons emerging in more realistic, $(4+d)$-dimensional, Kaluza–Klein theories, in which d dimensions are spontaneously compactified to form a space with topology T generated by an isometry group G (in the five-dimensional theory $T=S^1$, $G=U(1)$). The solitons in the theory will be then potentially classified by the possible $T(S^1)$ bundles over S^2 at spacelike infinity. This $(2+d)$-dimensional manifold might then be a boundary of a $(3+d)$-dimensional manifold which would represent the constant time sections of a static soliton solution. Based on our five-dimensional evidence one then has reason to hope that solutions with this topology actually exist, as well as a strategy for constructing them.

For example one might consider a 6-dimensional theory, compactified so that T is the 2 sphere S^2, and the isometry group G is SO(3). Such a Kaluza–Klein vacuum can be constructed, following Freund and Rubin [21] and Randjbar-Daemi, Salam and Strathdee [22], by adding an abelian gauge field for the six-dimensional gravitational theory. Since there are only two kinds of S^2 bundles over S^2 (a direct product and a twisted product) one therefore expects to find Z_2 monopoles in such a theory. From the four-dimensional point of view this is a SO(3) Yang–Mills theory coupled to gravity, and again one might expect to find $Z_2 = \pi^1(SO(3))$ monopoles. It would be interesting to construct these, since in this theory there are no massless dilatons, and the strange gravitational properties of our five-dimensional monopoles should be absent. Even more interesting, of course, would be to search for solitons in a realistic version of Kaluza–Klein theory, which produced correct low-energy physics.

We would like to thank a number of our colleagues for very useful discussions, especially Edward Witten and Sidney Coleman.

Note added in proof:

Recently, Sorkin [23] has also found the magnetic monopole solution in the Kaluza–Klein theory. We would like to thank the referee for pointing out some related work of Pollard [24], who discussed spherically symmetric static solutions of the five-dimensional Einstein equations.

References

[1] T. Kaluza, Sitzungsber Preuss. Akad. Wiss. Phys. Math. K1 (1961) 996;
O. Klein, Z. Phys. 37 (1926) 895

[2] A. Einstein and P. Bergmann, Ann. of Math. 39 (1938) 683;
P. Jordan, Ann. of Phys. (Leipzig) 1 (1947) 219;
Y. Thiry, C.R. Acad. Sci. 226 (1948) 216; 226 (1948) 1881;
A. Lichnerowicz, Théories relativistes de la gravitation et de l'electromagnetisme, A. Masson et Cie, Paris (1955)

[3] O. Klein, New theories in physics (Martinus Nijshoff, den Haag, 1939);
B.S. deWitt, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965);
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Reports Math. Phys. 1 (1970) 291;
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1972) 1711;
Y.M. Cho, J. Math. Phys. 16 (1975) 2029

[4] E. Cremmer and J.H. Schwarz, Phys. Lett. 57B (1975) 463; Nucl. Phys. B108 (1976) 409; B118 (1977) 61

[5] E. Witten, Nucl. Phys. B186 (1981) 412

[6] A. Einstein and W. Pauli, Ann. of Math. 44 (1943) 131

[7] A. Einstein and J. Grommer, Scripta Univ. Atque Bibl. Hierosolymitanarum I (1923) 1

[8] E. Witten, Nucl. Phys. B195 (1982) 481

[9] A. Chodos and T. Appelquist, Phys. Rev. Lett. 50 (1983) 141

[10] C. Brans and R.H. Dicke, Phys. Rev. 124 (1961) 95

[11] R.H. Dicke, Phys. Rev. 125 (1962) 2163

[12] G. 't Hooft, Nucl. Phys. B79 (1974) 276;
A. Polyakov, ZhetF Pisma Red. 20 (1974) 430

[13] E.T. Newman, L. Tamburino and T. Unti, J. Math. Phys. 4 (1963) 915;
S.W. Hawking, Phys. Lett. 60A (1977) 81

[14] C.W. Misner, J. Math. Phys. 4 (1963) 924

[15] R. Dashen, B. Hasslacher and A. Neveu, Phys. Rev. D10 (1974) 4114, 4130; D11 (1975) 3424;
J. Goldstone and R. Jackiw, Phys. Rev. D11 (1975) 1486;
J.L. Gervais and B. Sakita, Phys. Rev. D11 (1975) 2943;
J.L. Gervais, A. Jevicki and B. Sakita, Phys. Rev. D12 (1975) 1038;
C. Callan and D. Gross, Nucl. Phys. B93 (1975) 29

[16] G.W. Gibbons and S.W. Hawking, Phys. Lett. 78B (1978) 430

[17] G.W. Gibbons and S.W. Hawking, Comm. Math. Phys. 66 (1979) 291

[18] R.P. Kerr, Phys. Rev. Lett. 11 (1963) 237;
G.W. Gibbons and S.W. Hawking, Phys. Rev. D15 (1977) 2752

[19] V.A. Belinskii and V.E. Zakharov, JETP (Sov. Phys.) 48 (1978) 985

[20] G.W. Gibbons and M.J. Perry, Phys. Rev. D22 (1980) 313

[21] P. Freund and M. Rubin, Phys. Lett. 97B (1980) 233

[22] S. Randjbar-Daemi, A. Salam and J. Strathdee, Nucl. Phys. B214 (1983), 491

[23] R. Sorkin, Phys. Rev. Lett. 51 (1983) 87

[24] D. Pollard, J. Phys. A16 (1983) 565

NON-ABELIAN KALUZA–KLEIN MONOPOLES

Malcolm J. PERRY[1]
Department of Physics, Princeton University, NJ 08544, USA

Received 16 December 1983

I classify magnetic monopoles in general Kaluza–Klein theories. The construction of monopole solutions is explicitly carried out for an SO(3) model.

Kaluza–Klein theories [1] provide a promising approach to the unification of gravitation with other interactions. Part of the motivation for their consideration is that supergravity theories are naturally formulated in spacetime dimensions greater than four. Even if supergravity is not the ultimate answer to quantum gravity, superstring theory is a promising candidate [2]. It can only be consistently formulated in ten spacetime dimensions. Whatever the ultimate description of quantum gravity is, if that theory is based on spacetime dimensions greater than four, the remaining dimensions must be compactified in order to provide a phenomenologically satisfactory theory.

In a typical Kaluza–Klein theory, one starts with the Einstein action, perhaps supplemented by an auxiliary gauge field [3] or with the supersymmetric extension of the Einstein action [4]. Sometimes an auxiliary gauge field is required by supersymmetry, as can be seen by the example of $N = 1$ supergravity in 11 spacetime dimensions. One supposes that the $(d + 4)$-dimensional spacetime is locally of the form $\mathrm{M} \times \Sigma$, where Σ is some compact d-dimensional manifold. The spatial extent of Σ is assumed to be very small, typically the Planck length. This ensures that Σ cannot be directly observed. Σ however does manifest itself at low energies indirectly by the appearance of vector fields which resemble Yang–Mills fields with a gauge group G. G is the isometry group of Σ, which is presumed to act effectively on Σ. There will also be a set of scalar fields which transform covariantly under the action of G.

Suppose that $(d + 4)$-dimensional spacetime has a set of coordinates x^A $(A, B \ldots = 1, 2, \ldots, d + 4)$. These are divided into x^a $(a, b \ldots = 1, 2, 3, 4)$ which lie in M, a Lorentz-signatured four-dimensional spacetime which represents "ordinary" spacetime. x^α $(\alpha, \beta \ldots = 5, \ldots, d + 4)$ are a set of coordinates that lie in Σ, the compact d-dimensional space. Σ has a set of n $[d \leqslant n \leqslant \frac{1}{2}d(d + 1)]$ global Killing vectors K^i $(i, j \ldots = 1, \ldots, n)$ which have the commutation relations

$$[K^i, K^j] = -c^{ijk} K^k, \tag{1}$$

c^{ijk} are the structure constants of the isometry group G. Since Σ has global Killing vectors, it follows that Σ must have positive scalar curvature [5].

By an appropriate choice of Σ, G can be made into almost any gauge group.

Several examples are easy to construct from homogeneous spaces. If $\Sigma = \mathrm{S}^n$, then $\mathrm{G} = \mathrm{SO}(n + 1)$; if $\Sigma = \mathrm{CP}^n$ then $\mathrm{G} = \mathrm{SU}(n + 1)$; if $\Sigma = \mathrm{QP}^n$ then $\mathrm{G} = \mathrm{Sp}(n + 1)$; if $\Sigma = \mathrm{HP}^2$ then $\mathrm{G} = \mathrm{F}^4$. A more realistic set of possible Σ's has been provided by Witten [1], an example of which is $\Sigma = \mathrm{S}^5 \times \mathrm{S}^3/\mathrm{S}^1$ when $\mathrm{G} = \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$.

The set of n vector gauge fields is given by $A^i_a = g_{a\alpha} K^{i\alpha}$ and the $\frac{1}{2}n(n+1)$ scalars by $\phi^{ij} = g_{\alpha\beta} K^{i\alpha} K^{j\beta}$ [‡1].

---

[1] Research supported by the National Science Foundation under grant PHY80-19754, and the Alfred P. Sloan Foundation.

[‡1] It should be noted that gauge fields have been normalized so that the gauge coupling constant does not explicitly appear in the definition of either A^i_a, or the action.

The gauge transformations of A and ϕ correspond to coordinate transformations restricted to M, reflecting the fact that Diff^{d+4} has been spontaneously broken to $\mathrm{Diff}^4 \times \mathrm{G}$.

One would like to know if there is any way of distinguishing this type of unification from purely four-dimensional approaches by low-energy observations. One possibility is to look at the topological excitations in the theory. In a conventional grand unified theory in which the GUT group G′ is spontaneously broken down to H, there are monopoles which are classified by $\pi^1(\mathrm{H})$, or equivalently since H is a subgroup of G′, $\pi^2(\mathrm{G}'/\mathrm{H})$ [6]. In Kaluza–Klein theories, monopoles arise because the fibration of Σ over M can have fixed points. If we suppose that M is asymptotically flat, then a classification of the gauge fields is provided by examination of their behaviour at spatial infinity. Thus we classify the topological excitations which look like monopoles by asking what the topological structure of spacelike infinity is. At spacelike infinity, the spacetime will be a G-valued bundle over S^2. The topologically distinct bundles are thus classified by $\pi^1(\mathrm{G})$ [7]. We can list $\pi^1(\mathrm{G})$ for the classical groups G. For $\mathrm{G} = \mathrm{U}(n)$, $\pi^1(\mathrm{G}) = \mathrm{Z}$; for $\mathrm{G} = \mathrm{SO}(n)$, $\pi^1(\mathrm{G}) = \mathrm{Z}_2$; and for $\mathrm{G} = \mathrm{Sp}(n), \mathrm{SU}(n), \mathrm{E}_6, \mathrm{E}_7, \mathrm{E}_8, \mathrm{F}_4$ or G_2, $\pi^1(\mathrm{G}) = 0$ [8]. In ref. [9], it was shown how to construct magnetic monopoles in the abelian version of the Kaluza–Klein theory. In the remainder of this paper, we explicitly show how to construct monopoles in an SO(3) Kaluza–Klein theory. The generalization to SO(n) is probably straightforward.

We are interested specifically in the case where G = SO(3). Such a gauge group can be obtained by letting Σ be S^2 with a metric of constant curvature. The $(d+4)$-metric restricted to this S^2 is, up to a conformal factor which is independent of x^α, given by

$$ds_\Sigma^2\ d\psi^2 + \sin^2\psi\, d\chi^2 . \tag{2}$$

$x^\alpha = (\psi, \chi^0)$ are the standard polar coordinates on a sphere. This space has a set of three Killing vectors K^i which obey the commutation relations (1) with the structure constants being those of SO(3). Explicitly

$$K^1 = -\sin\chi \frac{\partial}{\partial\psi} - \cot\psi \cos\chi \frac{\partial}{\partial\chi},$$

$$K^2 = \cos\chi \frac{\partial}{\partial\psi} - \cot\psi \sin\chi \frac{\partial}{\partial\chi}, \quad K^3 = \frac{\partial}{\partial\chi}. \tag{3}$$

We shall choose for convenience a gauge such that $A^1 = A^2 = 0$, and restrict ourselves to solutions which are spherically symmetric in their gravitational field. The six-dimensional metric can then be written as

$$ds^2 = -h^2\, dt^2 + g^2\, dr^2 + k^2(d\theta^2 + \sin^2\theta\, d\phi^2)$$
$$+ p^2\,[d\psi^2 + \sin^2\psi\,(d\chi + q\cos\theta\, d\phi - s\, dt)^2], \tag{4}$$

g, h, k, p, q and s can all be functions only of r. In this case

$$A^3 = p(q\cos\theta\, d\phi - s\, dt). \tag{5}$$

It thus represents both a magnetic field and an electric field unless q or s vanish. Furthermore, unless q is constant, with $q = \frac{1}{2}n$ (n an integer), there will be string singularities along the rotation axes of $\partial/\partial\chi$, namely $\psi = 0$ (north axis) or $\psi = \pi$ (south axis). If $n = 0 \bmod 2$ then this is a metric on a space which at spatial infinity is the direct product of S^2 over S^2 and hence topologically trivial. If $n = 1 \bmod 2$ then at spatial infinity, the space is not homeomorphic to the product space but to the twisted S^2 bundle over S^2, which is topologically equivalent to $\mathrm{CP}^2 \# \mathrm{CP}^2$ [10]. This illustrates the possible Z_2 quantum number of these topological excitations. $n/2$ can be thought of as the monopole number.

In order to investigate the details of such topological objects, we must choose a specific theory. There are two basic ways to achieve a compactification which gives rise to some compact dimensions leaving others uncompactified. One is spontaneous compactification [2], achieved at the tree level. The other is compactification induced by quantum corrections [11]. Spontaneous compactification is most easily arranged by the introduction of an auxiliary antisymmetric tensor field which has k indices, and is derived from the covariant curl of a $k-1$ indexed antisymmetric potential

$$F_{a_1 \ldots a_k} = \nabla_{[a_1} A_{a_2 \ldots a_k]} . \tag{6}$$

k is chosen to be either d or 4, so that the separation of spacetime into d- and 4-dimensional sectors is natural. This type of field arises naturally in supergravity theories. The action can then be taken to be proportional to

$$\int \frac{-1}{16\pi}(R - 2\Lambda) + \frac{1}{2k} F_{a_1 \ldots a_k} F^{a_1 \ldots a_k} (\det g_{ab})^{1/2} \, \mathrm{d}^{d+4} x. \tag{7}$$

R is the Ricci scalar of the $(d + 4)$-dimensional space. Λ is the cosmological constant which must be included if the four-dimensional spacetime is to be asymptotically flat. Of course, one can include other fields including supersymmetric extensions if one wishes. The Einstein equations result from extremizing this action with respect to the metric.

We are specifically interested in the case of $d = 2$. Thus, if $k = 2$ the Einstein equations are

$$R_{ab} - \tfrac{1}{2} R g_{ab} + \Lambda g_{ab} = 8\pi (F_{ac} F_b{}^c - \tfrac{1}{4} g_{ab} F_{cd} F^{cd}). \tag{8}$$

If $k = 4$, the Einstein equations are

$$R_{ab} - \tfrac{1}{2} R g_{ab} + \Lambda g_{ab} = 8\pi (F_{acde} F_b{}^{cde} - \tfrac{1}{8} g_{ab} F_{cdef} F^{cdef}). \tag{9}$$

Somewhat remarkably there are no solutions to (8) with $q \neq 0$ within the context of the metric ansatz in eq. (4). Thus if we choose a two-indexed gauge field, we will not find magnetic monopoles in such a theory.

If we choose $k = 4$, we find solutions with the metric given by the ansatz (4). The only non-vanishing components of F_{abcd} are specified as follows:

$$F_{\mathrm{tr}\theta\phi} = (\Lambda/24\pi)^{1/2} r^2 \sin\theta,$$

$$F_{\mathrm{tr}\,\psi\chi} = q(192\pi\Lambda)^{-1/2} r^{-2} \sin\psi \cos\psi,$$

$$F_{\mathrm{tr}\,\psi\phi} = q^2(192\pi\Lambda)^{-1/2} r^{-2} \sin\psi \cos\psi \cos\theta,$$

$$F_{\theta\phi\psi\chi} = q(192\pi\Lambda)^{-1/2} \sin\psi \cos\psi \sin\theta,$$

$$F_{\theta\phi\psi t} = q^2(192\pi\Lambda)^{-1/2} \sin\psi \cos\psi \sin\theta \cos\theta. \tag{10}$$

The remaining components of F can be found by using its antisymmetry properties. The functions appearing in the spacetime metric are

$$k = r, \quad p = 1/\sqrt{2\Lambda}, \quad q = \text{const}, \quad s = \pm q/r,$$

$$h^2 = g^{-2}(1 - 2\mu/r + q^2/4\Lambda r^2). \tag{11}$$

μ is an arbitrary constant, and represents the mass of this object [‡2]. As previously noted, provided $q = n/2$ there are no string singularities along the axes $\psi = 0$ and $\psi = \pi$. The point $r = 0$ is a curvature singularity. However for fixed χ and ψ the geometry resembles that of the Reissner–Nordstrom spacetime.

In a coordinate system in which χ is replaced by χ such that

$$\mathrm{d}\chi = \mathrm{d}\chi + q \cos\theta \, \mathrm{d}\phi - s \, \mathrm{d}t \tag{12}$$

we can see that the geometry, for fixed ψ, χ, is precisely that of the Reissner–Nordstrom solution. (Note that $\partial/\partial\chi$ is still a Killing vector.) Thus we can interpret this solution as either a black hole if $\mu^2 \geqslant q^2/4\Lambda$, or a naked singularity otherwise. If $\mu^2 > q^2/4\Lambda$, then the spacetime is unstable due to Hawking effect [12]. Note that the black hole cannot radiate away its "charge" as it does in the four-dimensional Reissner–Nordstrom solution [13], because the charge in this case is topological in character. However if $\mu^2 = q^2/4\Lambda$, then the Hawking temperature associated with the horizon at

$$r = |q|/2\sqrt{\Lambda} \tag{13}$$

vanishes, and we have a stable object. In the six-dimensional Kaluza–Klein theory, this (with $q = \frac{1}{2}$) is the stable Z_2 monopole. It should be noted that if $q = 1$, the space is topologically equivalent to be vacuum. However the space is not flat since the assembly of two monopoles requires the concentration of energy at some point in the spacetime. It should be observed that this object has similarities to the gravitational solitons discussed by Hajicek [14], and Gibbons and Hull [15].

It should be noted that this solution possesses both an electric and a magnetic field. If we define the electric field by

$$E^i_a G^i_{ab} l^b$$

and the magnetic field by

$$B^i_a = \tfrac{1}{2} \epsilon_{abcd} G^{i\,bc} l^d,$$

where $G^i_{ab} = \nabla_a A^i_b - \nabla_b A^i_a$, ($\nabla_a$ and ϵ_{abcd} being with respect to the four-dimensional metric g_{ab}) and $l^a \partial/\partial x^a = \partial/\partial t$, then (with $q = \frac{1}{2}$)

[‡2] Unlike the abelian monopole, there is no apparent violation of the principle of equivalence associated with this solution [9].

$$E_r^3 \overset{r\to\infty}{=} 1/4\Lambda r^2, \quad B_r^3 \overset{r\to\infty}{=} 1/4\Lambda r^2.$$

Thus the magnetic charge is equal to the electric charge.

However, as was shown by Witten [16], if one lives in a world in which $\theta \neq 0$, then there are monopoles with electric charge $\theta/2\pi$, and magnetic charge 1/2. Thus, in the six-dimensional Kaluza–Klein theory, we discover that we are automatically living in a world in which $\theta = \pi$. It seems quite likely that if we drop the assumption of spherical symmetry in finding solutions of the field equations, we would be able to find dyonic excitations of this monopole with all possible odd half-integral electric charges.

The fact that $\theta = \pi$ is a topological fact which arises from the topology of the space at spatial infinity ($r \to \infty$ t = const), which is ($\mathrm{CP}^2 \# \overline{\mathrm{CP}^2}$). $\mathrm{CP}^2 \# \overline{\mathrm{CP}^2}$ does not admit a spin structure since its second Stieffel–Whitney class does not vanish. It does however admit a spin G like structure [17]. In a spin G structure there is an abnormal charge statistics relation in which bosons must have integral charges, and fermions add half-integral charges. The magnetic monopole behaves as a fermion, and hence compels $\theta = \pi$.

Finally, it should be noted that since the size of the Kaluza–Klein space determines the coupling constant for the gauge fields, fixing Λ also fixes the coupling constant. Thus there is really only a single arbitrary parameter, Λ, in the solutions presented here.

One expects that this construction should be easily generazible to theories in higher dimensions. However, since the only monopoles one can find in any Kaluza–Klein theory have either Z or Z_2 charges, one does not expect to find anything new in higher dimensions. Furthermore, the analysis given here is almost certainly easy to generalize to spacetimes whose non-compact dimensions are asymptotic to anti-de Sitter space. Neither matter will be pursued here.

I wish to thank Curt Callen for an interesting conversation, and Andy Strominger for his encouragement.

References

[1] T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. K1 (1921) 996;
O. Klein, Z. Phys. 37 (1926) 895;
B.S. de Witt, Dynamical theory of groups and fields (Gordon and Breach, New York, 1965);
Y.M. Cho and P.G.O. Freund, Phys. Rev. D12 (1972) 1711;
Y.M. Cho, J. Math. Phys. 16 (1975) 2029;
R. Kerner, Ann. Inst. H. Poincaré 9 (1968) 143;
A. Trautman, Rep. Math. Phys. 1 (1970) 291;
E. Witten, Nucl. Phys. B186 (1981) 412.
[2] J.H. Schwarz, Phys. Rep. 89 (1982) 223.
[3] P.G.O. Freund and M. Rubin, Phys. Lett. 97B (1980) 233.
[4] P. van Nieuwenhuizen, Phys. Rep. 68 (1981) 189;
E. Cremmer and B. Julia, Nucl. Phys. B159 (1979) 141.
[5] K. Yano and T. Nagano, Ann. Math. 69 (1959) 451.
[6] S. Coleman, The magnetic monopole fifty years later, Harvard preprint (1982).
[7] J. Feldbau, C.R. Acad. Sci. Paris 209 (1939) 1621.
[8] N. Steenrod, The topology of fibre bundles (Princeton U.P., Princeton, 1951).
[9] D.J. Gross nad M.J. Perry, Nucl. Phys. B226 (1951) 29.
[10] D.N. Page, Phys. Lett. 79B (1978) 235.
[11] A. Chodos and S. Detweiler, Phys. Rev. D21 (1980) 2167;
S. Weinberg and P. Candelas, to be published.
[12] S.W. Hawking, Commun. Math. Phys. 43 (1975) 199.
[13] G.W. Gibbons, Commun. Math. Phys. 44 (1975) 245.
[14] P. Hajicek, Nucl. Phys. B185 (1981) 254; Phys. Rev. D26 (1982) 3384, 3396.
[15] G.W. Gibbons and C.M. Hull, Phys. Lett. 109B (1982) 190.
[16] E. Witten, Phys. Lett. 114B (1982) 137.
[17] S.J. Avis and C.J. Isham, Commun. Math. Phys. 72 (1980) 103.

CHAPTER

10

Superstrings

SUPERSYMMETRICAL STRING THEORIES [☆]

Michael B. GREEN [1] and John H. SCHWARZ
California Institute of Technology, Pasadena, CA 91125, USA

Received 11 December 1981

A supersymmetrical light-cone-gauge string action is presented. It provides a basis for understanding the previously-studied supersymmetrical dual string theory as well as two new closed-string theories that have extended supersymmetry in ten dimensions, corresponding to $N = 8$ supersymmetry in four dimensions.

In several recent papers [1–3], we have developed a new oscillator description of the supersymmetrical dual string theory [4]. It has proved to be superior to older formulations for understanding the ten-dimensional supersymmetry of the theory and for explicitly calculating scattering amplitudes. This letter describes a light-cone-gauge string action from which this oscillator formalism can be derived. The string action also leads to the discovery of two new closed-string theories with extended supersymmetry in ten dimensions (corresponding to $N = 8$ supersymmetry in four dimensions). Unlike any other known dual string theory, these closed-string theories turn out to be completely free of divergences in the one-loop approximation (which is all that has been checked so far).

The conventional (Veneziano) string model can be based on an action principle for the string coordinates $X^{\mu}(\sigma, \tau)$ [5] [‡1]. The action can be written as a σ–τ reparametrization-invariant expression by using an auxiliary two-dimensional metric tensor. The local invariance allows one to choose an orthonormal gauge in which the light-cone component $X^{-}(\sigma, \tau)$ may be regarded as a dependent variable [7]. The remaining gauge freedom (plus partial use of equations of motion) allows one to set

$$X^{+}(\sigma, \tau) = x^{+} + 2\alpha' p^{+} \tau, \tag{1}$$

where x^{+} and p^{+} are constants. Once this is done, the string dynamics resides entirely in the transverse coordinates $X^{i}(\sigma, \tau)$. A description of the supersymmetrical string theories requires fermionic degrees of freedom, $S^{Aa}(\sigma, \tau)$, as well. The index $A = 1, 2$ describes a spinor in the σ–τ space, whereas the index a is 32-valued, as appropriate for a spinor in 10-dimensional spacetime. S^{1a} and S^{2a} are Majorana–Weyl spinors in ten dimensions. For much of the analysis it is not necessary to specify whether S^{1} and S^{2} have the same or opposite handedness, but ultimately both cases will turn out to be of interest. Corresponding to eq. (1), the fermionic field is required to satisfy the gauge condition

$$(\gamma^{+})^{ab} S^{Ab} = 0. \tag{2}$$

The postulated two-dimensional string action in the light-cone gauge is [‡2]

[☆] Work supported in part by the US Department of Energy under Contract No. DE-AC03-81-ER40050.

[1] On leave from Queen Mary College, University of London, England.

[‡1] As customary, σ is a spacelike variable that runs from 0 to π along the string, while τ is a timelike evolution parameter. These are represented collectively by a two-vector index α. μ is a vector index for the physical spacetime, 26-valued for the Veneziano model and 10-valued for the supersymmetrical models. For a collection of review articles, see ref. [6].

[‡2] For the two-dimensional Dirac matrices, the Majorana representation $\rho^{0} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\rho^{1} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$ is convenient. We also define $\bar{S}^{Aa} = S^{\dagger Bb} (\gamma^{0})^{ba} (\rho^{0})^{BA}$, $\rho_{3} = \rho^{0} \rho^{1} = \mathrm{diag}(1, -1)$. A spacelike metric is used for both two-vectors and ten-vectors. We henceforth set $\alpha' = 1/2$.

$$\int d\tau \int_0^\pi d\sigma [-(4\pi\alpha')^{-1} \partial_\alpha X^i \partial^\alpha X^i + (\mathrm{i}/4\pi) \bar{S} \gamma^- \rho^\alpha \partial_\alpha S] . \tag{3}$$

A covariant gauge-invariant action from which eqs. (1)–(3) can be deduced is not yet known. The equations of motion implied by (3), subject to boundary conditions to be discussed, are

$$\partial^\alpha \partial_\alpha X^i = 0 \quad \text{and} \quad \rho^\alpha \partial_\alpha S = 0. \tag{4}$$

This action is invariant under the supersymmetry transformations

$$\delta X^i = (p^+)^{-1/2} \bar{\epsilon} \gamma^i S, \tag{5a}$$

$$\delta S = \mathrm{i}(p^+)^{-1/2} \gamma_- \gamma_\mu (\rho \cdot \partial X^\mu) \epsilon, \tag{5b}$$

again assuming appropriate boundary conditions. These transformations may be contrasted with those in an earlier formulation in which ϵ was a spinor only in the two-dimensional sense [8]. The parameters ϵ^{Aa} are Majorana–Weyl in $d = 10$, and may either be regarded as parametrizing two supersymmetries in ten dimensions or 16 supersymmetries in two dimensions. They may even depend on σ and τ provided that

$$(\partial/\partial\tau - \rho_3 \partial/\partial\sigma)\epsilon = 0, \tag{6}$$

but we are primarily interested in the case of constant ϵ. Commuting two global supersymmetry transformations, one finds (using $\rho \cdot \partial S = 0$) that

$$[\delta_1, \delta_2] X^i = \xi^\alpha \partial_\alpha X^i + a^i, \qquad [\delta_1, \delta_2] S = \xi^\alpha \partial_\alpha S, \tag{7a,b}$$

with

$$\xi^\alpha = -(2\mathrm{i}/p^+) \bar{\epsilon}^{(1)} \gamma_- \rho^\alpha \epsilon^{(2)}, \tag{8}$$

$$a^i = -2\mathrm{i} \bar{\epsilon}^{(1)} \rho^0 \gamma^i \epsilon^{(2)}, \tag{9}$$

which may be interpreted as translations of (σ, τ) and X^i, respectively. The supersymmetry charges obtained from eqs. (3) and (5) by the Noether method are

$$Q^{Aa} = \frac{\mathrm{i}}{\pi} (p^+)^{-1/2} \int_0^\pi [\gamma_\mu (\rho \cdot \partial X^\mu) \rho^0 S]^{Aa} \mathrm{d}\sigma. \tag{10}$$

Boundary conditions needed to derive the equations of motion (4) must be considered carefully. In the case of open strings, it is necessary that S^1 and S^2 have the same handedness and

$$\partial X^i(0, \tau)/\partial\sigma = \partial X^i(\pi, \tau)/\partial\sigma = 0, \tag{11}$$

$$S^{1a}(0, \tau) = S^{2a}(0, \tau), \qquad S^{1a}(\pi, \tau) = S^{2a}(\pi, \tau). \tag{12a, b}$$

Possible minus signs in eq. (12) are excluded by requiring invariance under global supersymmetry transformations. In this case, one also needs $\epsilon^{1a} = -\epsilon^{2a}$, so that there is, in fact, just one ten-dimensional supersymmetry invariance in the open-string case. Eqs. (4), (11), (12) lead to the mode expansions

$$X^i = x^i + p^i \tau + \mathrm{i} \sum_{n \neq 0} \frac{1}{n} \alpha_n^i \cos n\sigma \, \mathrm{e}^{-\mathrm{i}n\tau}, \tag{13}$$

$$S^{1a} = \sum_{n=-\infty}^{\infty} S_n^a \exp[-\mathrm{i}n(\tau - \sigma)], \tag{14a}$$

$$S^{2a} = \sum_{n=-\infty}^{\infty} S_n^a \exp[-\mathrm{i}n(\tau + \sigma)], \tag{14b}$$

Substituting these expansions in eq. (10), one finds that Q^{1a} and Q^{2a} are not separately conserved (they depend on τ) but that

$$Q^a = \tfrac{1}{2}(Q^{1a} + Q^{2a})$$

$$= \mathrm{i}(p^+)^{1/2} (\gamma_+ S_0)^a + \mathrm{i}(p^+)^{-1/2} \sum_{-\infty}^{\infty} (\gamma_i S_{-n})^a \alpha_n^i \tag{15}$$

is conserved.

Canonical quantization gives

$$[\alpha_m^i, \alpha_n^j] = m \delta_{m+n,0} \delta^{ij}, \tag{16}$$

$$\{S_m^a, \bar{S}_n^b\} = (\gamma^+ h)^{ab} \delta_{m+n,0}, \tag{17}$$

$$[\alpha_m^i, S_n^a] = 0, \tag{18}$$

where h represents a Weyl projection operator $\tfrac{1}{2}(1 \pm \gamma_{11})$. Eqs. (15)–(18) were discovered in ref. [1], where it is proved that

$$\{Q^a, \bar{Q}^b\} = -2(h\gamma \cdot p)^{ab}, \tag{19}$$

provided that one imposes the mass-shell condition

$$p^- = (1/2p^+)(p^i p^i + N), \tag{20}$$

$$N = \sum_{n=1}^{\infty} (\alpha_{-n}^i \alpha_n^i + \tfrac{1}{2} n \bar{S}_{-n} \gamma^- S_n). \tag{21}$$

Ref. [1] gives formulas for the Lorentz generators $J^{\mu\nu}$ in terms of the α_n^i and the S_n^a and proves that they have the correct algebra when eq. (20) is satisfied. A proof that Q^a transforms as a spinor in the ten-dimensional sense is also given there. Lorentz covariance is not at all obvious in the light-cone-gauge formalism [the action in eq. (3) only has manifest covariance in the eight-dimensional transverse space]. However, functional integral methods analogous to those developed for the conventional model [9] can be used to obtain covariant S-matrix elements from the light-cone-gauge action. The results are the same as obtained by the operator methods in our previous works [1–3].

Turning attention to closed strings, we replace the boundary conditions in eqs. (11) and (12) with the requirement that X^i and S^{Aa} are periodic in σ with period π. This, together with eq. (4), leads to the closed-string mode expansions

$$X^i(\sigma, \tau) = x^i + p^i\tau + \frac{\mathrm{i}}{2}\sum_{n\neq 0}\frac{1}{n}\{\alpha^i_n \exp[-2\mathrm{i}n(\tau-\sigma)] + \tilde{\alpha}^i_n \exp[-2\mathrm{i}n(\tau+\sigma)]\}, \tag{22}$$

$$S^{1a} = \sum_{-\infty}^{\infty} S^a_n \exp[-2\mathrm{i}n(\tau-\sigma)], \tag{23a}$$

$$S^{2a} = \sum_{-\infty}^{\infty} \tilde{S}^a_n \exp[-2\mathrm{i}n(\tau+\sigma)]. \tag{23b}$$

The canonical commutation relations give eqs. (16)–(18) for the operators with tildes as well as the ones without. Also

$$[\alpha^i_m, \tilde{\alpha}^j_n] = [\alpha^i_m, \tilde{S}^a_n] = [\tilde{\alpha}^i_m, S^a_n] = \{S^a_m, \tilde{S}^b_n\} = 0. \tag{24}$$

The handedness of S^1 and S^2 may be independently assigned. Substituting the mode expansions (22) and (23) into eq. (10) gives

$$Q^{1a} = \mathrm{i}(p^+)^{1/2}(\gamma_+ S_0)^a + 2\mathrm{i}(p^+)^{-1/2}\sum_{-\infty}^{\infty}(\gamma_i S_{-n})^a \alpha^i_n, \tag{25a}$$

$$Q^{2a} = \mathrm{i}(p^+)^{1/2}(\gamma_+ \tilde{S}_0)^a + 2\mathrm{i}(p^+)^{-1/2}\sum_{-\infty}^{\infty}(\gamma_i \tilde{S}_{-n})^a \tilde{\alpha}^i_n. \tag{25b}$$

These are separately conserved and give the algebra of extended supersymmetry in ten dimensions

$$\{Q^{Aa}, \bar{Q}^{Bb}\} = -2(h^A\gamma\cdot p)^{ab}(\rho^0)^{AB}. \tag{26}$$

The mass-shell condition in eq. (20) is replaced for closed strings by the conditions

$$L_0 = \tfrac{1}{8}p^2 + N = 0, \qquad \tilde{L}_0 = \tfrac{1}{8}p^2 + \tilde{N} = 0, \tag{27a,b}$$

where N is given in eq. (21) and $\tilde{N}$ is the corresponding expression with $\tilde{\alpha}$ and $\tilde{S}$. (The closed-string Regge slope is $\frac{1}{4}$—half of the open-string slope.) Eq. (27) is required both for the closure of the Lorentz algebra and for the validity of the supersymmetry algebra in eq. (26). One can define a truncation to a closed-string theory with simple supersymmetry by restricting the physical spectrum to states that are symmetrical under the interchanges $\alpha^i_m \leftrightarrow \tilde{\alpha}^i_m$ and $S^a_m \leftrightarrow \tilde{S}^a_m$. (This requires that S^a_m and $\tilde{S}^a_m$ have the same handedness, $h^1 = h^2$.) In this case $Q = \frac{1}{2}(Q^1 + Q^2)$ is the only remaining supersymmetry. This truncated theory is of importance because it describes the closed strings that interact with the open strings describes above, whereas untruncated closed-string models cannot interact with open strings (just as $N = 4$ Yang–Mills cannot couple to $N = 8$ supergravity).

The 256 massless physical closed-string states may be described by ket vectors $|ij\rangle$, $|ia\rangle$, $|ai\rangle$, $|ab\rangle$ where the first label refers to the S_0 space and the second to $\tilde{S}_0$ space.

$$|ai\rangle = \tfrac{1}{8}\mathrm{i}(\gamma_j S_0)^a|ji\rangle, \qquad |ia\rangle = \tfrac{1}{8}\mathrm{i}(\gamma_j \tilde{S}_0)^a|ij\rangle, \tag{28a,b}$$

$$|ab\rangle = -\tfrac{1}{64}(\gamma_i S_0)^a(\gamma_j \tilde{S}_0)^b|ij\rangle. \tag{28c}$$

The supersymmetry and Lorentz transformation rules for these states can be deduced from eqs. (17) and (25), as well as the Lorentz generator representations, which have the form given in ref. [1] with identical extra terms formed from oscillators with tildes. The spin content of this multiplet is precisely that of extended supergravity in ten dimensions. There are two distinct cases. If the S and $\tilde{S}$ oscillators have opposite handedness then the two gravitinos contained in $|ai\rangle$ and $|ia\rangle$ have opposite handedness and the states $|ab\rangle$ correspond to a vector and third-rank antisymmetric tensor field. This corresponds to the spin content obtained by reduction of eleven-dimensional supergravity to ten dimensions. If S and $\tilde{S}$ have the same handedness, then the two gravitinos have the same handedness and the states $|ab\rangle$ correspond to a scalar, second-rank antisymmetric tensor, and fourth-rank antisymmetric self-dual tensor. This is clearly a distinct alternative and cannot be obtained by reduction from a higher dimension. Nonetheless, the two theories are extremely close in their structure. It is amusing that even though the spin content of the two theories is different in ten dimensions, it coincides on reduction to any lower dimension.

The interacting closed-string theories can be described in the operator formalism in terms of a propagator for states satisfying eq. (27):

$$\Delta = \int_{|z|<1} \mathrm{d}^2 z\, z^{L_0 - 1}\, \bar{z}^{\tilde{L}_0 - 1}, \tag{29}$$

and vertices describing the emission of massless states. The vertex formulas are given as direct products of open-string vertices [2] in the tilde and no tilde spaces. More specifically, representing the open-string vertex for emitting a massless vector with polarization vector ζ^μ as $\zeta^\mu V^B_\mu$ and the vertex for emitting a massless fermion with Dirac spinor u^a as $\bar{u}^a V^\Gamma_a$. the closed string vertices have the form $\zeta^{\mu\nu} V^B_\mu \tilde{V}^B_\nu$, $\bar{u}^{\mu a}_1 V^B_\mu \tilde{V}^\Gamma_a$, $\bar{u}^{a\mu}_2 V^\Gamma_a \tilde{V}^B_\mu$, and $\bar{\eta}^{ab} V^\Gamma_a \tilde{V}^\Gamma_b$. Using the vertex formulas given in ref. [2], tree and loop amplitudes for specific processes are straightforward to calculate. One finds, for example, that one-loop graphs with three or fewer external (on mass-shell) lines are identically zero for the same reasons as in the open-string case [3], namely the vanishing of zero-mode traces.

The one-loop contribution to the N-particle connected S-matrix is given by a single Feynman graph in the case of closed strings. For the case of any four massless external states, the one-loop contribution consists of a kinematical factor involving the polarization vectors and spinors as well as the momenta of the external states times the integral

$$\int \frac{\mathrm{d}^2\tau}{(\mathrm{Im}\,\tau)^5} \int \prod_{i=1}^{3} \mathrm{d}^2\nu_i \prod_{1\leqslant i<j\leqslant 4} \{\exp[-\pi(\mathrm{Im}\,\nu_{ji})^2/\mathrm{Im}\,\tau] \times |\vartheta_1(\nu_{ji}|\tau)|\}^{(1/2)k_i\cdot k_j}. \tag{30}$$

In this integral ϑ_1 is a Jacobi elliptic function, $\nu_{ji} = \nu_j - \nu_i$ and $\nu_4 = \tau$. The ν_i integration regions are restricted by

$$0 \leqslant \mathrm{Im}\,\nu_i \leqslant \mathrm{Im}\,\tau, \qquad -\tfrac{1}{2} \leqslant \mathrm{Re}\,\nu_i \leqslant \tfrac{1}{2}. \tag{31}$$

To determine the appropriate τ integration region, one notes that eq. (30) may be written in the form $\int \mathrm{d}^2\tau\,(\mathrm{Im}\,\tau)^{-2} F(\tau)$, where F is an automorphic function. That is, it is invariant under the modular group

$$\tau \to (a\tau + b)/(c\tau + d), \tag{32}$$

where a, b, c and d are integers satisfying $ad - bc = 1$. Therefore, to avoid multiple counting, one must restrict the τ integration to the "fundamental" region [10]

$$-\tfrac{1}{2} \leqslant \mathrm{Re}\,\tau \leqslant \tfrac{1}{2}, \qquad \mathrm{Im}\,\tau \geqslant 0, \qquad |\tau| \geqslant 1. \tag{33}$$

It is easy to see that the one-loop integral with these limits has the expected poles and branch cuts and is otherwise completely free from divergences. This contrasts, in particular, with the conventional Shapiro–Virasoro model in which there is a divergence associated with the corner where $\nu_1, \nu_2, \nu_3 \to \tau$ [10]. A massless scalar particle at zero momentum in the propagator connecting a tadpole to a tree graph is responsible for the divergence. This does not occur in the supersymmetrical theory, because the supersymmetry algebra makes loops with three or fewer on-shell massless states vanish identically.

In conclusion, we have found a string action that explains the oscillator description of the theory of open and closed strings with simple supersymmetry in ten dimensions. It also leads to the discovery of two new closed-string theories, both of which correspond to $N = 8$ supergravity in four dimensions in the limit that the radii of six compactified dimensions and the slope parameter α' are shrunk to zero. If it should turn out that $N = 8$ supergravity is not finite to all orders in perturbation theory, but that these closed-string theories are, then they may provide the means to make sense of the quantum theory.

We are grateful to Lars Brink for helpful discussions.

References

[1] M.B. Green and J.H. Schwarz, Nucl. Phys. B181 (1981) 502.
[2] M.B. Green and J.H. Schwarz, Caltech preprint CALT-68-872.
[3] M.B. Green and J.H. Schwarz, Caltech preprint CALT-68-873.
[4] A. Neveu and J.H. Schwarz, Nucl. Phys. B31 (1971) 86; Phys. Rev. D4 (1971) 1109;
P. Ramond, Phys. Rev. D3 (1971) 2415.
[5] Y. Nambu, Proc. Intern. Conf. on Symmetries and quark models (Wayne State University, 1979) (Gordon and Breach, New York, 1970), p. 269;
L. Susskind, Nuovo Cimento 69A (1970) 457.

ANOMALY CANCELLATIONS IN SUPERSYMMETRIC $D = 10$ GAUGE THEORY AND SUPERSTRING THEORY ☆

Michael B. GREEN
Queen Mary College, University of London, London E1 4NS, UK
and California Institute of Technology, Pasadena, CA 91125, USA

and

John H. SCHWARZ
California Institute of Technology, Pasadena, CA 91125, USA

Received 10 September 1984

Supersymmetric ten-dimensional Yang–Mills theory coupled to $N = 1$, $D = 10$ supergravity has gauge and gravitational anomalies that can be partially cancelled by the addition of suitable local interactions. The remaining pieces of all the anomalies cancel if the gauge group is SO(32) or $E_8 \times E_8$. These cancellations are automatically incorporated in the type I superstring theory based on SO(32). A superstring theory for $E_8 \times E_8$ has not yet been constructed.

It is not easy to obtain an effective four-dimensional theory containing chiral fermions by spontaneous compactification of a theory in $D > 4$ dimensions. It may be necessary that the D-dimensional theory itself contains chiral fermions and elementary Yang–Mills gauge fields [1]. But such theories are, in general, plagued by anomalies in the divergences of Yang–Mills gauge currents [2] and gravitational (local Lorentz) gauge currents [3]. Certain superstring theories are promising candidates for ten-dimensional quantum theories containing gravitation. (For reviews see ref. [4].) An essential requirement for their consistency is the absence of gauge and gravitational anomalies.

An interesting example of an anomaly-free chiral theory in higher dimensions is IIB $D = 10$ supergravity [5], which can be viewed as the low-energy approximation to the corresponding superstring theory (based on oriented closed strings). Gravitational anomalies have been shown to cancel in IIB supergravity [3], and probably do so in IIB superstring theory, as well. Since these theories contain no elementary Yang–Mills gauge fields, they are trivially free from Yang–Mills gauge anomalies. Unfortunately, the absence of such gauge fields may also exclude compactifications leading to a chiral theory in four dimensions [‡1].

In this letter we wish to report the discovery of other anomaly-free chiral theories in ten dimensions. One of these is type I superstring theory (which contains unoriented open and closed strings) with SO(32) as the gauge group. The mechanisms responsible for anomaly cancellations are explained below in more conventional field-theoretic terms by describing the low-energy effective field theory, which consists of $N = 1$, $D = 10$ supergravity coupled to supersymmetric Yang–Mills theory, even though any truncation of the complete string theory is probably not fully consistent. At the level of the effective field theory, $E_8 \times E_8$ is shown to be another possible anomaly-free theory.

The superstring results are summarized at the end of this paper, but details of the analysis are left for a future publication. The coupled $D = 10$ super Yang–

☆ Work supported in part by the US Department of Energy under Contract No. DEAC-03-81ER400050 and by the Fleischmann Foundation.

[‡1] An interesting suggestion for circumventing this problem has been made recently [6].

Mills plus supergravity theory in its minimal form [7] contains gauge and gravitational anomalies. In order to understand the cancellations, the minimal lagrangian must be supplemented by additional higher-derivative local interactions that arise in the low-energy expansion of the superstring theory. Some of these terms are noninvariant under local Lorentz and Yang–Mills gauge transformations, with variations arranged to cancel those arising from fermion loops. They are analogous to "Wess–Zumino terms" that have been introduced in other contexts. Although such terms can be added to the action for any gauge group, for SO(32) or $E_8 \times E_8$ they can be chosen to completely cancel the anomalies. From the point of view of the effective field theory this appears to involve arbitrary choices of coefficients. However, these are precisely the coefficients that are contained in the SO(32) superstring theory.

The action for the interacting theory contains the bosonic terms

$$S_0 = \int \mathrm{d}^{10}x\, e[-(1/2\kappa^2)R - (1/\kappa^2)\varphi^{-2}\partial_\mu\varphi\partial^\mu\varphi - (1/4g^2)\varphi^{-1}F^a_{\mu\nu}F^{\mu\nu a} - (3\kappa^2/2g^4)\varphi^{-2}H_{\mu\nu\rho}H^{\mu\nu\rho}]\,. \tag{1}$$

This appears to contain two coupling constants, the Yang–Mills coupling g and the gravitational coupling κ, with dimensions of $(\text{length})^3$ and $(\text{length})^4$, respectively. (The ratio κ/g^2 is proportional to the string tension.) However, g^2 is not an arbitrary parameter of the action since it can be absorbed in the definition of φ. Thus there are no arbitrary dimensionless parameters. (We are grateful to E. Witten for pointing this out to us.) In eq. (1) e is the determinant of the zehnbein e_μ^m, and φ is a scalar field. The Yang–Mills field strength is defined (in the language of forms) by

$$F \equiv \tfrac{1}{2}F_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu = \mathrm{d}A + A^2\,. \tag{2}$$

We usually omit $\wedge$ signs for simplicity, so that $A^2 \equiv A \wedge A$. The one-form potential A is a matrix representation of the gauge algebra

$$A = A^a_\mu \lambda^a \mathrm{d}x^\mu\,. \tag{3}$$

We only discuss SO(N), USp(N) and $E_8 \times E_8$ explicitly, but other groups can be analyzed in an analogous manner. It is important for the analysis that the λ matrices are antihermitean matrices in the *fundamental* representation of the algebra (the N of SO(N) or USp(N)), even though the massless fermions of the Yang–Mills sector belong to the adjoint representation The trace of a matrix M is denoted by tr M when M is in the fundamental representation and by Tr M when M is in the adjoint representation. (Of course, these are identical for E_8.) The field strength H associated with the two-form potential $B = B_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$ of the supergravity multiplet is defined for the case of SO(N) or USp(N) by

$$H = \mathrm{d}B + \omega^0_{3\mathrm{L}} - \omega^0_{3\mathrm{L}}\,. \tag{4}$$

In the case of $E_8 \times E_8$ this equation is replaced by

$$H = \mathrm{d}B + \tfrac{1}{30}\omega^0_{3\mathrm{Y}_1} + \tfrac{1}{30}\omega^0_{3\mathrm{Y}_2} - \omega^0_{3\mathrm{L}}\,. \tag{4$'$}$$

In eq. (4) $\omega^0_{3\mathrm{Y}}$ is the Yang–Mills Chern–Simons three-form

$$\omega^0_{3\mathrm{Y}} = \mathrm{tr}(AF - \tfrac{1}{3}A^3)\,, \tag{5}$$

which satisfies

$$\mathrm{d}\omega^0_{3\mathrm{Y}} = \mathrm{tr}\,F^2\,. \tag{6}$$

In eq. (4′) $\omega^0_{3\mathrm{Y}_1}$ and $\omega^0_{3\mathrm{Y}_2}$ are Yang–Mills Chern–Simons forms (given by eq. (5)) for each of the two E_8 factors. The Yang–Mills Chern–Simons terms in eq. (4) would have a coefficient of $\frac{1}{30}$ (as in eq. (4′)) for the special case of SO(32) if it were defined in terms of matrices in the adjoint representation. The Lorentz group Chern–Simons form is given in identical fashion by

$$\omega^0_{3\mathrm{L}} = \mathrm{tr}(\omega R - \tfrac{1}{3}\omega^3) \tag{7}$$

and satisfies

$$\mathrm{d}\omega^0_{3\mathrm{L}} = \mathrm{tr}\,R^2\,. \tag{8}$$

Here $\omega = \omega_\mu \mathrm{d}x^\mu$, the local Lorentz connection, is a 10×10 matrix of the tangent-space algebra SO(9,1), and $R = \mathrm{d}\omega + \omega^2$ is the Lorentz-curvature two-form. The last term in eq. (4) is a higher-derivative term arising in the low-energy expansion of the superstring and therefore does not appear in ref. [7]. Its inclusion leads to a very similar treatment of the Yang–Mills local Lorentz symmetries.

Under infinitesimal local Yang–Mills transformations (with parameter Λ) and local Lorentz transformations (with parameter Θ), the fields transform as follows

$$\delta A = \mathrm{d}\Lambda + [A, \Lambda]\,, \quad \delta\omega = \mathrm{d}\Theta + [\omega, \Theta]\,, \tag{9a,b}$$

$$\delta B = -\operatorname{tr}(A\,\mathrm{d}\Lambda) + \operatorname{tr}(\omega\,\mathrm{d}\Theta)\,,$$

$$\text{for } \mathrm{SO}(N) \quad \text{or } \mathrm{USp}(N)\,, \tag{9c}$$

$$\delta B = -\tfrac{1}{30}\operatorname{tr}_1(A_1\mathrm{d}\Lambda_1) - \tfrac{1}{30}\operatorname{tr}_2(A_2\mathrm{d}\Lambda_2) + \operatorname{tr}(\omega\,\mathrm{d}\Theta)\,,$$

$$\text{for } \mathrm{E}_8 \times \mathrm{E}_8\,, \tag{9c$'$}$$

$$\delta F = [F, \Lambda]\,, \quad \delta R = [R, \Theta]\,, \quad \delta H = 0\,. \tag{10a,b,c}$$

The subscripts in eq. (9c′) refer to the two E_8 factors. It follows from eq. (10) that the terms in eq. (1) are invariant. It is crucial that the potential B transforms under both kinds of transformations and that the field strength H is invariant.

Cancellation of Yang–Mills gauge anomalies. Anomalies in ten-dimensional gauge theories arise from loop diagrams with six or more external lines and chiral fields circulating around the loop. The anomalies corresponding to diagrams with more than six external lines are uniquely determined by the hexagon diagrams. Their form can be deduced by implementing the Wess–Zumino consistency conditions [8]. The analysis of ref. [9] relates the anomalies to a formal expression in twelve dimensions, namely the gauge-invariant twelve-form Ω_{12}, where

$$\Omega_{4n} = \operatorname{tr} F^{2n}\,. \tag{11}$$

The cases of $\mathrm{SO}(N)$ and $\mathrm{USp}(N)$ are considered first. In these cases Ω_{4n} can be reexpressed in terms of fundamental-representation matrices

$$\Omega_4 = (N + 2l)\operatorname{tr} F^2\,, \tag{12a}$$

$$\Omega_8 = (N + 8l)\operatorname{tr} F^4 + 3(\operatorname{tr} F^2)^2\,, \tag{12b}$$

$$\Omega_{12} = (N + 32l)\operatorname{tr} F^6 + 15\operatorname{tr} F^2 \operatorname{tr} F^4\,, \tag{12c}$$

where

$$l = -1 \quad \text{for } \mathrm{SO}(N)\,, \tag{13a}$$

$$l = +1 \quad \text{for } \mathrm{USp}(N)\,. \tag{13b}$$

The first term in eq. (12c) corresponds to anomalies that cannot be cancelled by the addition of local interactions to the effective action. However, it is absent if the gauge group is chosen to be SO(32). The second term in eq. (12c) can be expressed as

$$\operatorname{tr} F^2 \operatorname{tr} F^4 = \mathrm{d}(\omega^0_{3\mathrm{Y}} \operatorname{tr} F^4) = \mathrm{d}(\omega^0_{7\mathrm{Y}} \operatorname{tr} F^2)\,, \tag{14}$$

where $\omega^0_{2n+1\,\mathrm{Y}}$ is defined (up to an exact form) by

$$\mathrm{d}\omega^0_{2n+1\,\mathrm{Y}} = \operatorname{tr} F^{n+1}\,. \tag{15}$$

Explicit expressions for $\omega^0_{2n+1\,\mathrm{Y}}$ are given in ref. [9]. Similarly, applying an infinitesimal gauge transformation, one defines $\omega^1_{2n\,\mathrm{Y}}$ (up to an exact form) by

$$\delta\omega^0_{2n+1\,\mathrm{Y}} = -\mathrm{d}\omega^1_{2n\,\mathrm{Y}}\,, \tag{16}$$

In the present case, the consistent anomaly corresponding to the second term in eq. (12c) is

$$G = c\int \left(\tfrac{1}{3}\omega^1_{2\mathrm{Y}} \operatorname{tr} F^4 + \tfrac{2}{3}\omega^1_{6\,\mathrm{Y}} \operatorname{tr} F^2\right), \tag{17}$$

where the ratio of the two terms is uniquely determined by Bose symmetry. The numerical constant. c can be deduced from refs. [3,9]. This anomaly can be cancelled by adding to the effective action

$$S_1 = c\int \left(B \operatorname{tr} F^4 + \tfrac{2}{3}\omega^0_{3\mathrm{Y}}\omega^0_{7\mathrm{Y}}\right). \tag{18}$$

The nontrivial Λ transformation of B in eq. (9c) is the key to the cancellation between δS_1 and G. This amounts to a complete cancellation of the Yang–Mills anomaly for the gauge group SO(32).

The extra terms in the action contribute anomalous pieces to tree amplitudes. For example, there is a tree diagram with six external gauge fields (fig. 1) in which the $B \operatorname{tr} F^4$ vertex is connected to the $\omega^0_{3\mathrm{Y}}\ \mathrm{d}B$ vertex (contained in the H^2 term of eq. (1)) by a B propagator. This gives rise to an anomalous piece of the six-particle amplitude, which together with the second term in eq. (18) cancels the hexagon anomaly.

For the group E_8, eq. (12) is replaced by

$$\Omega_4 = \operatorname{tr} F^2 \tag{12a$'$}$$

$$\Omega_8 = \tfrac{1}{100}(\operatorname{tr} F^2)^2 \tag{12b$'$}$$

$$\Omega_{12} = \tfrac{1}{7200}(\operatorname{tr} F^2)^3\,. \tag{12c$'$}$$

It is important in comparing the analysis of the mixed

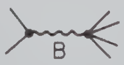


Fig. 1. Six-particle tree amplitude that contributes to anomalies.

anomalies for SO(32) and $E_8 \times E_8$ that for both these gauge groups

$$\Omega_{12} = \tfrac{1}{48}\Omega_4[\Omega_8 - \tfrac{1}{300}(\Omega_4)^2] \,. \tag{19}$$

It follows from eq. (12c′) by the same reasoning as before that the consistent anomaly for $E_8 \times E_8$ is given by

$$G' = \frac{c}{108\,000}\int[\omega^1_{2Y_1}(\mathrm{tr}_1 F^2)^2 + \omega^1_{2Y_2}(\mathrm{tr}_2 F^2)^2] \,. \tag{17′}$$

This can be cancelled by adding

$$S'_1 = \frac{c}{108\,000}\int\{30B[(\mathrm{tr}_1 F^2)^2 + (\mathrm{tr}_2 F^2)^2 - \mathrm{tr}_1 F^2\,\mathrm{tr}_2 F^2] - \omega^0_{3Y_1}\omega^0_{3Y_2}(\mathrm{tr}_1 F^2 - \mathrm{tr}_2 F^2)\} \,. \tag{18′}$$

Cancellation of gravitational anomalies. Gravitational anomalies can be viewed either as a breakdown of general coordinate invariance or local Lorentz invariance [10,11]. They can be shifted from the one to the other by adding a suitable extra term to the effective action [11]. This term is zero when either anomaly is absent. We choose to study local Lorentz anomalies, rather than general-coordinate ones, since they bear the greater similarity to the Yang–Mills gauge ones. As described above for the Yang–Mills case, Lorentz anomaly structures arising from hexagon (and higher) diagrams with external gravitons only can be associated with the formal twelve-form [3]

$$-[\tfrac{1}{32} + (n-496)/13824]\,(\mathrm{tr}\,R^2)^3 - [\tfrac{1}{8} + (n-496)/5760]\,\mathrm{tr}\,R^2\,\mathrm{tr}\,R^4 - [(n-496)/7560]\,\mathrm{tr}\,R^6 . \tag{20}$$

In this expression we have included the contributions of one left-handed spin 3/2 gravitino and one right-handed spin 1/2 field from the supergravity sector and n left-handed spin 1/2 fields from the matter sector, which only depends on the dimension of the gauge group.

The last term in eq. (20) corresponds to an anomaly of the form $\int\omega^1_{10L}$, which cannot be cancelled by adding local terms to the action. Therefore 496 left-handed spin 1/2 fields are needed in the matter sector in order that it vanish. Remarkably, since the dimension of the adjoint representation of SO(32) or $E_8 \times E_8$ is 496, the cancellation occurs for either of these gauge groups. The anomalies associated with the first two terms of eq, (20) can be cancelled (putting $n = 496$) by adding to the effective action for SO(32) or $E_8 \times E_8$

$$S_2 = -c\int[\tfrac{1}{32}B(\mathrm{tr}\,R^2)^2 + \tfrac{1}{8}B\,\mathrm{tr}\,R^4 + \tfrac{1}{12}\omega^0_{3L}\omega^0_{7L}] \,. \tag{21}$$

Thus, not only does the cancellation of the Yang–Mills anomaly require coupling to the supergravity multiplet, but also the cancellation of the gravitational anomalies requires the contributions of Yang–Mills supermultiplets.

Cancellation of mixed anomalies. The extra terms in the action, which are needed to cancel the pure Yang–Mills and local Lorentz anomalies, also contribute to mixed anomalies. (Consider, for example, the Lorentz variation of the B field in eq. (18)). The relative normalization of all the anomalies can be deduced from eq. (63) of ref. [3]. In particular, the twelve-form that characterizes the mixed anomalies for SO(32) is

$$-\mathrm{tr}\,R^2[\mathrm{tr}\,F^4 + \tfrac{1}{8}(\mathrm{tr}\,F^2)^2] + \mathrm{tr}\,F^2[\tfrac{1}{8}\mathrm{tr}\,R^4 + \tfrac{5}{32}(\mathrm{tr}\,R^2)^2] \,. \tag{22}$$

The twelve-forms in (14), (20) and (22) are given with the correct relative normalization. The mixed anomalies can be cancelled by adding the local terms

$$S_3 = c\int(\tfrac{1}{8}B\,\mathrm{tr}\,R^2\,\mathrm{tr}\,F^2 + \tfrac{1}{48}\omega^0_{3L}\omega^0_{3Y}\,\mathrm{tr}\,R^2 - \tfrac{1}{24}\omega^0_{3Y}\omega^0_{3L}\,\mathrm{tr}\,F^2 - \tfrac{2}{3}\omega^0_{3L}\omega^0_{7Y} + \tfrac{1}{12}\omega^0_{3Y}\omega^0_{7L}) \,. \tag{23}$$

The cancellation requires that $N + 8l = 24$ and $N + 2l = 30$ in eq. (12). Thus the cancellation of mixed anomalies gives two independent confirmations of the consistency of choosing the gauge group SO(32). In summary, modifying the action of ref. [7] by including the last term in eq. (4) and adding $S_1 + S_2 + S_3$ to the action gives complete cancellation of all anomalies at one-loop level for the gauge group SO(32).

Given the cancellation of mixed anomalies in SO(32) the possibility of their cancellation in $E_8 \times E_8$ follows from eq. (19). Explicitly, for $E_8 \times E_8$ one obtains

$$S_3' = \frac{c}{7200} \int (30\, B \operatorname{tr} R^2 \operatorname{tr}_1 F^2 + 5\omega_{3\mathrm{L}}^0 \omega_{3\mathrm{Y}_1}^0 \operatorname{tr} R^2$$
$$+ 20\omega_{3\mathrm{Y}_1}^0 \omega_{7\mathrm{L}}^0$$
$$+ \omega_{3\mathrm{Y}_1}^0 \omega_{3\mathrm{L}}^0 \operatorname{tr}_1 F^2 + \omega_{3\mathrm{L}}^0 \omega_{3\mathrm{Y}_1}^0 \operatorname{tr}_2 F^2) + (1 \leftrightarrow 2). \tag{23'}$$

Thus, as before, the addition of $S_1' + S_2 + S_3'$ to the action cancels all the one-loop anomalies for $\mathrm{E}_8 \times \mathrm{E}_8$. It may be interesting to look for analogous anomaly-free theories in other dimensions.

Superstring anomalies. The results described above were discovered by calculating the anomalies of superstring diagrams and studying the low-energy expansions of the formulae. Since type I superstrings appear otherwise to be consistent quantum theories, the possibility of anomalies is of crucial importance. Their occurrence in a covariant formalism implies that unphysical modes of gauge fields cannot be consistently decoupled, leading to a breakdown of unitarity in the quantum theory. In the light-cone gauge formalism the problem would be expressed as a breakdown of Lorentz invariance.

The groups SO(N), USp(N), and U(N) have been shown to be possible superstring gauge groups at the classical level [4,12]. The U(N) case was shown in ref. [4] to give inconsistencies at the quantum level. In view of the preceding results, it seems very likely to us that $\mathrm{E}_8 \times \mathrm{E}_8$ is also possible for superstrings. Since such a superstring theory has not yet been formulated, the following discussion is restricted to the cases SO(N) and USp(N).

We have explicitly evaluated the anomalies of one-loop superstring hexagons with external massless Yang–Mills particles in a covariant formalism. The results are summarized here, but the details will be presented elsewhere. The one-loop diagrams come in three topologies, shown in fig. 2, called planar, nonorientable, and nonplanar. The planar and nonorientable diagrams have a group theory structure corresponding to the tr F^6 term in eq. (12) whereas the nonplanar diagrams have the group-theoretic structure of the tr F^2 tr F^4 term. The nonplanar diagrams turn out to be anomaly-free. This means that the superstring theory necessarily contains terms such as in eq. (18) with the requisite coefficients.

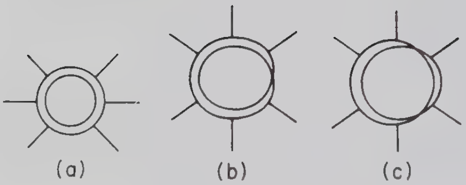


Fig. 2. (a) The planar hexagon string diagram. (b) The nonorientable hexagon string diagram. (c) The nonplanar hexagon string diagram.

The planar and nonorientable superstring diagrams do have anomalies. They are evaluated by associating polarization vectors ζ_r^μ to five of the external lines (r = 2, 3, 4, 5, 6)

$$k_r \cdot \zeta_r = k_r^2 = 0\,, \tag{24}$$

and the momentum vector k_1^μ in place of ζ_1^μ for line no. 1. Just as in ordinary field-theory calculations, various infinities must be carefully regulated in order to get the correct nonzero result. The result has the form

$$G = (\mathrm{const.})\,(N + 32l) \operatorname{tr}(\lambda_1 \lambda_2 \lambda_3 \lambda_4 \lambda_5 \lambda_6)$$
$$\times\, \epsilon_{\mu_1 \ldots \mu_5 \nu_1 \ldots \nu_5} \zeta_1^{\mu_1} \ldots \zeta_5^{\mu_5} k_1^{\nu_1} \ldots k_5^{\nu_5}$$
$$\times \int_0^1 \left(\prod_{i=1}^{5} \vartheta(\nu_{i+1} - \nu_i) \mathrm{d}\nu_i \right)$$
$$\times \prod_{1 \leqslant i < j \leqslant 8} [\sin \pi(\nu_j - \nu_i)]^{2\alpha' k_i \cdot k_j}$$
$$+ \text{ permutations}\,, \tag{25}$$

where $\nu_6 = 1$. The coefficient N comes from the planar diagram and the coefficient $32l$ from the nonorientable one. As before, a perfect cancellation occurs for SO(32). Eq. (25) is different from the anomaly of the low-energy effective field theory, because (contrary to common belief) massive fermions also contribute to the anomaly. It reduces to the usual anomaly at low energies ($\alpha' k_i \cdot k_j \ll 1$). Anomalies involving excited external states can be deduced by factorizing on poles of eq. (25).

The field-theoretic discussion in the earlier part of this paper only displayed certain higher-derivative terms, occurring in the low-energy expansion of the superstring, that were required for explaining the anomaly cancellation mechanisms. Clearly, only adding these terms and no others destroys supersymmetry. We suspect that restoring it would require introducing the entire infinite expansion of the superstring effective action. We have not investigated anomalous divergences of the supersymmetry current, but since the gravitino and the graviton are in the same supermultiplet it seems likely that the cancellation of gravitational anomalies should also ensure the cancellation of supersymmetry anomalies in the superstring theory.

In conclusion, it appears very likely that SO(32) superstring theory is a consistent quantum theory. More analysis of anomaly cancellation in various diagrams involving closed strings still needs to be carried out. Also, the possibility of anomalies associated with "big" gauge transformations, not connected to the identity, remains to be investigated. The anomaly cancellations in the case of the $E_8 \times E_8$ effective theory suggest that it is also the low-energy expansion of a superstring theory, which has not yet been formulated. The fact that anomaly-free type I superstring theories are chiral theories containing elementary gauge fields makes them appealing candidates for a fundamental theory.

We wish to acknowledge helpful and stimulating discussions with W. Bardeen, D. Friedan, S. Shenker, P. West, B. Zumino, and other participants at the Aspen Center for Physics. In an earlier version of this paper it was noted that the gravitational anomalies cancel for $E_8 \times E_8$, but the cancellation of the other anomalies in that theory was overlooked. We are grateful to G. Chapline and R. Slansky for pointing out to us that E_8 has no independent fourth- and sixth-order Casimir invariants. L. Dixon, J. Harvey and E. Witten and, independently, J. Thierry-Mieg have also observed that all the Yang–Mills and mixed anomalies can be cancelled for the case of $E_8 \times E_8$.

References

[1] E. Witten, Fermion numbers in Kaluza–Klein theory, Princeton University preprint.
[2] P.H. Frampton and T.W. Kephart, Phys. Rev. Lett. 50 (1983) 1343, 1347;
P.K. Townsend and G. Sierra, Nucl. Phys. B222 (1983) 493;
B. Zumino, Y.S. Wu and A. Zee, Nucl. Phys. B239 (1984) 477;
L. Baulieu, Nucl. Phys. B241 (1984) 557.
[3] L. Alvarez-Gaumé and E. Witten, Nucl. Phys. B234 (1983) 269.
[4] J.H. Schwarz, Phys. Rep. 89 (1982) 223; Caltech preprint CALT-68-1137;
M.B. Green, Surv. High Energy Phys. 3 (1983) 127.
[5] M.B. Green and J.H. Schwarz, Phys. Lett. 122B (1983) 143;
J.H. Schwarz and P.C. West, Phys. Lett. 126B (1983) 301;
J.H. Schwarz, Nucl. Phys. B226 (1983) 269;
P.Howe and P.C. West, Nucl. Phys. B238 (1984) 181.
[6] M. Gell-Mann and B. Zwiebach, Phys. Lett. 147B (1984) 111.
[7] E. Bergshoeff, M. de Roo, B. de Wit and P. van Nieuwenhuizen, Nucl. Phys. B195 (1982) 97;
G.F. Chapline and N.S. Manton, Phys. Lett. 120B (1983) 105.
[8] J. Wess and B. Zumino, Phys. Lett. 37B (1971) 95.
[9] B. Zumino, Y.S. Wu and A. Zee, Nucl. Phys. B239 (1984) 477.
[10] Z. Zumino, Les Houches lectures (1983);
R. Stora, Cargese lectures (1983).
[11] W.A. Bardeen and B. Zumino, Berkeley preprint LBL-17639;
L. Alvarez-Gaumé and P. Ginsparg, Harvard University preprint HUTP-84/AO16.
[12] N. Marcus and A. Sagnotti, Phys. Lett. 119B (1982) 97.,

CREDITS

CHAPTER II.

GUNNAR NORDSTRÖM, "ON THE POSSIBILITY OF A UNIFICATION OF THE ELECTROMAGNETIC AND GRAVITATION FIELDS." TRANSLATED FROM THE GERMAN ORIGINAL BY PETER G. O . FREUND.

GUNNAR NORDSTRÖM, "ÜBER DIE MÖGLICHKEIT, DAS ELEKTROMAGNETISCHE FELD UND DAS GRAVITATIONSFELD ZU VEREINIGEN," PHYSIK. ZEITSCHR. *15,* 504-506. (1914).

TH. KALUZA, "ON THE UNITY PROBLEM OF PHYSICS." THIS IS A MODIFIED VERSION (BY PETER G. O. FREUND) OF EARLIER TRANSLATIONS BY T. MUTA AND C. HOENSELAERS.

TH. KALUZA, "ZUM UNITÄTSPROBLEM DER PHYSIK," SITZUNGSBER. D. PREUSS. AKADEMIE D. WISSENSCHAFTEN, PHYSIK.-MATHEMAT. KLASSE, 966-972 (1921).

O. KLEIN, "QUANTUM THEORY AND FIVE DIMENSIONAL THEORY OF RELATIVITY," TRANS. C. HOENSELAERS, IN *UNIFIED FIELD THEORIES OF MORE THAN FOUR DIMENSIONS INCLUDING EXACT SOLUTIONS,* ED. V. DE SABBATA AND E. SCHMUTZER (SINGAPORE: WORLD SCIENTIFIC, 1983), 434-446.

O. KLEIN, "THE ATOMICITY OF ELECTRICITY AS A QUANTUM THEORY LAW," NATURE *118,* 516 (1926).

A. EINSTEIN, P. BERGMANN, "ON A GENERALIZATION OF KALUZA'S THEORY OF ELECTRICITY," ANN. OF MATH. *39,* 683-701 (1938).

Y. THIRY, "THE EQUATIONS OF KALUZA'S UNIFIED THEORY." TRANSLATION FROM THE FRENCH ORIGINAL BY PETER G. O. FREUND.

Y. THIRY, "LES ÉQUATIONS DE LA THÉORIE UNITAIRE DE KALUZA," C.R. ACAD. SCI. (PARIS) *226,* 216-218 (1948).

CHAPTER III.

B. S. DEWITT, "DYNAMICAL THEORY OF GROUPS AND FIELDS," PROBLEM #77, IN "RELATIVITY GROUPS AND TOPOLOGY," EDITED BY C. DEWITT AND B. DEWITT (GORDON AND BREACH, NEW YORK 1964) P. 725.

RYSZARD KERNER, "GENERALIZATION OF THE KALUZA-KLEIN THEORY FOR AN ARBITRARY NON-ABELIAN GAUGE GROUP," ANN. INST. HENRI POINCARÉ, *9,* 143-152 (1968).

Y. M. CHO, P. G. O. FREUND, "NON-ABELIAN GAUGE FIELDS AS NAMBU-GOLDSTONE FIELDS," PHYS. REV. *D12,* 1711-1720 (1975). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY.

E. CREMMER, J. SCHERK, "SPONTANEOUS COMPACTIFICATION OF SPACE IN AN EINSTEIN-YANG-MILLS-HIGGS MODEL," NUCLEAR PHYSICS *B108,* 409-416 (1976). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

J. F. LUCIANI, "SPACE-TIME GEOMETRY AND SYMMETRY BREAKING," NUCLEAR PHYSICS *B135,* 111-130 (1978). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

ABDUS SALAM, J. STRATHDEE, "ON KALUZA-KLEIN THEORY," ANNALS OF PHYSICS *141,* 316-352 (1982). ACADEMIC PRESS INC., NEW YORK.

CHAPTER IV.
E. CREMMER, B. JULIA, J. SCHERK, "SUPERGRAVITY THEORY IN 11 DIMENSIONS," PHYS. LETT. *76B,* 409-412 (1978). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

E. CREMMER, B. JULIA, "THE N = 8 SUPERGRAVITY THEORY. I. THE LAGRANGIAN," PHYS. LETT. *80B,* 48-51 (1978). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

P. G. O. FREUND, MARK A. RUBIN, "DYNAMICS OF DIMENSIONAL REDUCTION," PHYS. LETT. *97B,* 233-235 (1980). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

F. ENGLERT, "SPONTANEOUS COMPACTIFICATION OF ELEVEN-DIMENSIONAL SUPERGRAVITY," PHYS. LETT. *119B,* 339-342 (1982). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

I. BARS, S. W. MACDOWELL, "GRAVITY WITH EXTRA GAUGE SYMMETRY," PHYS. LETT. *129B,* 182-184 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

B. BIRAN, F. ENGLERT, B. DE WIT, H. NICOLAI, "GAUGED N = 8 SUPERGRAVITY AND ITS BREAKING FROM SPONTANEOUS COMPACTIFICATION," PHYS. LETT. *124B,* 45-50 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

A. CASHER, F. ENGLERT, H. NICOLAI, M. ROOMAN, "THE MASS SPECTRUM OF SUPERGRAVITY ON THE ROUND SEVEN-SPHERE," NUCLEAR PHYSICS *B243,* 173-188 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

M. A. AWADA, M. J. DUFF, C. N. POPE, "$N = 8$ SUPERGRAVITY BREAKS DOWN TO $N = 1$," PHYS. REV. LETT. *50,* 294-297 (1983). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY.

M. J. DUFF, B. E. W. NILSSON, C. N. POPE, "SPONTANEOUS SUPERSYMMETRY BREAKING BY THE SQUASHED SEVEN-SPHERE," PHYS. REV. LETT. *50,* 2043-2046 (1983) AND PHYS. REV. LETT. *51,* 846(E) (1983). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY.

R. D'AURIA, P. FRÉ, P. VAN NIEUWENHUIZEN, "SYMMETRY BREAKING IN $D = 11$ SUPERGRAVITY ON THE PARALLELIZED SEVEN-SPHERE," PHYS. LETT. *122B,* 225-231 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

FEZA GÜRSEY, CHIA-HSIUNG TZE, "OCTONIONIC TORSION ON S^7 AND ENGLERT'S COMPACTIFICATION OF $D = 11$ SUPERGRAVITY," PHYS. LETT. *127B,* 191-196 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

G. F. CHAPLINE, N. S. MANTON, "UNIFICATION OF YANG-MILLS THEORY AND SUPERGRAVITY IN TEN DIMENSIONS," PHYS. LETT. *120B,* 105-109 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM. ✓

D. Z. FREEDMAN, G. W. GIBBONS, P. C. WEST, "TEN INTO FOUR WON'T GO," PHYS. LETT. *124B,* 491-492 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

B. DE WIT, H. NICOLAI, "A NEW SO(7) INVARIANT SOLUTION OF $D = 11$ SUPERGRAVITY," PHYS. LETT. *148B,* 60-64 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

CHAPTER V.

EDWARD WITTEN, "SEARCH FOR A REALISTIC KALUZA-KLEIN THEORY," NUCLEAR PHYSICS *B186,* 412-428 (1981). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM. ✓

CHAPTER VI.

ALAN CHODOS, STEVEN DETWEILER, "WHERE HAS THE FIFTH DIMENSION GONE?," PHYS. REV. *D21,* 2167-2170 (1980). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY. ✓

P. G. O. FREUND, "KALUZA-KLEIN COSMOLOGIES," NUCLEAR PHYSICS *B209,* 146-156 (1982). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

ENRIQUE ALVAREZ, M. BELEN GAVELA, "ENTROPY FROM EXTRA DIMENSIONS," PHYS. REV. LETT. *51,* 931-934 (1983). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY.

Q. SHAFI, C. WETTERICH, "COSMOLOGY FROM HIGHER-DIMENSIONAL GRAVITY," PHYS. LETT. *129B,* 387-391 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

DESHDEEP SAHDEV, "TOWARDS A RELATIVISTIC KALUZA-KLEIN COSMOLOGY," PHYS. LETT. *137B,* 155-159 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

EDWARD W. KOLB, RICHARD SLANSKY, "DIMENSIONAL REDUCTION IN THE EARLY UNIVERSE: WHERE HAVE THE MASSIVE PARTICLES GONE?," PHYS. LETT. *135B,* 378-382 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

KEI-ICHI MAEDA, "IS COSMOLOGICAL DIMENSIONAL REDUCTION POSSIBLE?," PHYS. LETT. *138B,* 269-273 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

S. RANDJBAR-DAEMI, ABDUS SALAM, J. STRATHDEE, "ON KALUZA-KLEIN COSMOLOGY," PHYS. LETT. *13B,* 388-392 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

CHAPTER VII.

H. B. G. CASIMIR, "ON THE ATTRACTION BETWEEN TWO PERFECTLY CONDUCTING PLATES." REPRINTED WITH PERMISSION FROM PROCEEDINGS OF THE ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES VOL. 51 (1948) PP. 793-795.

THOMAS APPELQUIST, ALAN CHODOS, "QUANTUM DYNAMICS OF KALUZA-KLEIN THEORIES," PHYS. REV. *D28,* 772-784 (1983). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY.

STEVEN WEINBERG, "CHARGES FROM EXTRA DIMENSIONS," PHYS. LETT. *125B,* 265-269 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

MARK A. RUBIN, BERNARD D. BARKAN-ROTH, "TEMPERATURE EFFECTS IN FIVE-DIMENSIONAL KALUZA-KLEIN THEORY," NUCLEAR PHYSICS *B226,* 444-454 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

PHILIP CANDELAS, STEVEN WEINBERG, "CALCULATION OF GAUGE COUPLINGS AND COMPACT CIRCUMFERENCES FROM SELF-CONSISTENT DIMENSIONAL REDUCTION," NUCLEAR PHYSICS *B237,* 397-441 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

N. A. VORONOV, YA. I. KOGAN, "SPONTANEOUS COMPACTION IN KALUZA-KLEIN MODELS AND THE CASIMIR EFFECT," PIS'MA ZH. EKSP. TEOR. FIZ. *38,* 262-265 (1983); TRANSLATED AS *JETP LETTERS 38,* NO. 5, 311-314 (1983). © 1983 AMERICAN INSTITUTE OF PHYSICS.

G. W. GIBBONS, H. NICOLAI, "ONE-LOOP EFFECTS ON THE ROUND SEVEN-SPHERE," PHYS. LETT. *143B,* 108-114 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

T. INAMI, K. YAMAGISHI, "VANISHING QUANTUM VACUUM ENERGY IN ELEVEN-DIMENSIONAL SUPERGRAVITY ON THE ROUND SEVEN-SPHERE," PHYS. LETT. *143B,* 115-120 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

CHAPTER VIII.

EDWARD WITTEN, "FERMION QUANTUM NUMBERS IN KALUZA-KLEIN THEORY," PRINCETON PREPRINT (1983). AS APPEARED IN THE PROCEEDINGS OF THE SHELTER ISLAND CONFERENCE. THE MIT PRESS.

C. WETTERICH, "DIMENSIONAL REDUCTION OF FERMIONS IN GENERALIZED GRAVITY," NUCLEAR PHYSICS *B242,* 473-502 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

S. RANDJBAR-DAEMI, ABDUS SALAM, J. STRATHDEE, "SPONTANEOUS COMPACTIFICATION IN SIX-DIMENSIONAL EINSTEIN-MAXWELL THEORY," NUCLEAR PHYSICS *B214,* 491-512 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

STEVEN WEINBERG, "QUASI-RIEMANNIAN THEORIES OF GRAVITATION IN MORE THAN FOUR DIMENSIONS," PHYS. LETT. *138B,* 47-51 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

M. GELL-MANN, B. ZWIEBACH, "CURLING UP TWO SPATIAL DIMENSIONS WITH SU (1,1)/U(1)," PHYS. LETT. *147B,* 111-114 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

CHAPTER IX.

RAFAEL D. SORKIN, "KALUZA-KLEIN MONOPOLE," PHYS. REV. LETT. *51,* 87-90 (1983). REPRINTED BY PERMISSION OF THE AMERICAN PHYSICAL SOCIETY.

DAVID J. GROSS, MALCOLM J. PERRY, "MAGNETIC MONOPOLES IN KALUZA-KLEIN THEORIES," NUCLEAR PHYSICS *B226,* 29-48 (1983). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

MALCOLM J. PERRY, "NON-ABELIAN KALUZA-KLEIN MONOPOLES," PHYS. LETT. *137B,* 171-174 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

CHAPTER X.

M. B. GREEN, J. H. SCHWARZ, "SUPERSYMMETRICAL STRING THEORIES," PHYS. LETT. *109B,* 444-448 (1982). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

M. B. GREEN, J. H. SCHWARZ, "ANOMALY CANCELLATIONS IN SUPERSYMMETRIC $D=10$ GAUGE THEORY AND SUPERSTRING THEORY," PHYS. LETT. *149B,* 117-122 (1984). NORTH-HOLLAND PHYSICS PUBLISHING, AMSTERDAM.

⑩

The Addison-Wesley **Advanced Book Program** would like to offer you the opportunity to learn about our new physics titles in advance. To be placed on our mailing list and receive pre-publication notices and discounts, just **fill out this card completely** and return to us, postage paid. Thank you.

Name____

Title____

School/Company________________________________

Department____________________________________

Street Address__________________________________

City____________________State________Zip________

Telephone (________) ________________________

Where did you buy this book?

☐ Bookstore (non-campus)
☐ Mail Order
☐ School/Required for Class
☐ Campus Bookstore/Individual Study
☐ Toll Free # to Publisher
☐ Professional Meeting
☐ Other____

What professional physics & science associations are you an *active* member of?

☐ AAPT (American Association of Physics Teachers)
☐ AIP (American Institute of Physics)
☐ APS (American Physical Society)
☐ Sigma Pi Sigma
☐ SPS (Society of Physics Students)
☐ Other____

Check your areas of interest.

11 ☐ Quantum Mechanics
12 ☐ Particle/Astro Physics
13 ☐ Condensed Matter/Solid State Physics
14 ☐ Mathematical Physics
15 ☐ Nuclear Physics
16 ☐ Electron & Atomic Physics
17 ☐ Plasma Physics
18 ☐ Materials Physics
19 ☐ Biological Physics
20 ☐ High Polymer Physics
21 ☐ Chemical Physics
22 ☐ Fluid Dynamics
23 ☐ History of Physics
24 ☐ Statistical Physics
25 ☐ Other____

Are you more interested in: ☐ theory ☐ experimentation?

Are you currently writing or planning to write a physics book?

☐ Yes ☐ No
Area:____

(If Yes) **Are you interested in discussing your project with us?**

☐ Yes ☐ No

App

MAILING LIST • MAILING LIST • MAILING LIST • MAILING LIST • MAILING LIST • MAILING LIST

No Postage
Necessary
if Mailed in the
United States

BUSINESS REPLY CARD
FIRST CLASS PERMIT NO. 922 MENLO PARK, CA 94025

Postage will be paid by Addressee:

Addison-Wesley
Publishing Company, Inc.®
Advanced Book Program

2725 Sand Hill Road
Menlo Park, California 94025-9919

DATE DUE

FEB 19 '88
MAR 11 '88
4/5/88
~~4/25/88~~
5/31/88
6/22/88 DISCHARGED
DISCHARGED
SEP 08 '88
DISCHARGED
10/28/88
MAR 10 1997
APR 04 1997
REC'D APR 11 1997

DEMCO NO. 38-298

[redacted]

Modern Kaluza-Klein Theories

FRONTIERS IN PHYSICS SERIES (Volume No. 65)

First introduced in the 1920's, Kaluza-Klein theories have taken on a new and robust life in recent years. The advent of gauge field theories and ideas about grand unification have given new importance and immediacy to the further step of unifying elementary particle interactions with gravity. The Kaluza-Klein notion of extra dimensions offers perhaps the most elegant path to this tantalizing goal. This new reference volume is of particular and timely interest to particle physicists and astrophysicists, as it:

- Provides a comprehensive and up-to-date collection of the important articles since (and including) the original work of Kaluza and Klein, with emphasis on the work of the last decade;
- Covers all the major areas of application, with detailed introductions by the editors;
- Includes an extensive list of references.

About the Editors

Thomas Appelquist is a Professor of Physics and Department Chair at Yale University. In 1985 he was awarded Senior U.S. Scientist from the Alexander von Humboldt Foundation. Currently, Dr. Appelquist serves on the National Science Foundation Advisory Committee for Physics. After receiving his Ph.D. from Cornell University, he was an Assistant and Associate Professor of Physics at Harvard University. Professor Appelquist has been a Visiting Member at the Institute for Advanced Study at Princeton University, a Visiting Physicist at Fermi National Accelerator Laboratory, and has held fellowships from the Alfred P. Sloan Foundation and the American Physical Society.

Alan Chodos, a Senior Research Physicist and Lecturer at Yale University, received his Ph.D. in Theoretical Physics from Cornell University. Dr. Chodos was a Research Associate of the High Energy Theory Group at the University of Pennsylvania, and a Research Associate at the Center for Theoretical Physics at M.I.T.

Peter G.O. Freund is a Professor of Physics at the University of Chicago, where he has been since 1965. His work in the last decade has focused on super-symmetry and higher dimensional unification. Dr. Freund received his Ph.D. in Physics from the University of Vienna, and later became Research Associate at the Institute for Theoretical Physics in Vienna, a Chef de Travaux at the Institute for Theoretical Physics in Geneva, and a Research Associate and Assistant Professor at the Enrico Fermi Institute. Dr. Freund is a Fellow of the American Physical Society.

ISBN 0-201-09829-6/Hardbound/620 pp./1987

Books of Related Interest

Statistical Field Theory

Giorgio Parisi, Universita de Roma II, Tor Vergata, Frontiers in Physics Series (Volume No. 66), 0-201-05985-1, Hardbound, 608pp, 1987

Gauge Field Theories

Paul H. Frampton, University of North Carolina at Chapel Hill, Frontiers in Physics Series (Volume No. 62), 0-8053-2584-0, hardbound, 608pp, 1986.

Addison-Wesley Publishing Company, Inc.
The Advanced Book Program
Menlo Park, California • Reading, Massachusetts
Don Mills, Ontario • Wokingham, U.K. • Amsterdam • Sydney
Singapore • Tokyo • Madrid • Bogota • Santiago • San Juan

ISBN 0-201-09829-6